ercim-news.ercim.eu

Number 123 October 2020

ERCIM NEWS

Special theme:

Blue Growth

Also in this issue

Research and Innovation:

Evaluation of Synthetic Data for Privacy-Preserving Machine Learning

Editorial Information

ERCIM News is the magazine of ERCIM. Published quarterly, it reports on joint actions of the ERCIM partners, and aims to reflect the contribution made by ERCIM to the European Community in Information Technology and Applied Mathematics. Through short articles and news items, it provides a forum for the exchange of information between the institutes and also with the wider scientific community. This issue has a circulation of about 6,000 printed copies and is also available online.

ERCIM News is published by ERCIM EEIG
BP 93, F-06902 Sophia Antipolis Cedex, France
+33 4 9238 5010, contact@ercim.eu
Director: Philipp Hoschka, ISSN 0926-4981

Contributions
Contributions should be submitted to the local editor of your country

Advertising
For current advertising rates and conditions, see https://ercim-news.ercim.eu/ or contact peter.kunz@ercim.eu

ERCIM News online edition: *ercim-news.ercim.eu/*

Next issue:
January 2021: Epidemic Modelling

Subscription
Subscribe to ERCIM News by sending an email to en-subscriptions@ercim.eu

Cover photo by Luis Vidal on Unsplash

JOINT ERCIM ACTIONS

SPECIAL THEME

The special theme “Blue Growth” has been coordinated by the guest editors Alberto Gotta (ISTI-CNR) and John Mardikis (EPLO Circular Clima Institute)

Introduction to the special theme

RESEARCH AND INNOVATION

ANNOUNCEMENTS

ERCIM Workshop on Quality in AI

by Björn Levin (RISE) and Peter Kunz (ERCIM Office)

ERCIM invited experts in the field of artificial intelligence to a workshop held on 26 May 2020, with the aim of collecting ideas and suggestions about approaches to ensure high quality artificial intelligence (AI) in practical applications.

We briefly summarise the main items and ideas discussed in the workshop. The full report will be published on the ERCIM website. The discussed items have been grouped into five clusters:
- Governance – what factors and regulations should society consider to ensure the best use of AI, who should be responsible for what, and how should rules and regulations be developed?
- Trust – how is trust best built both among the public, and among professionals in other fields that use AI?
- Skills – how do we ensure that we have sufficiently skilled and responsible AI engineers?
- Process – how do we avoid repeating errors, how do we build and communicate best practices, and what public resources are needed?
- Testing – what can be tested, and what about the variables that are outside the realm of testing?
- Quality of data and methods – transparency vs integrity and trade secrets, deterioration over time, and the pros and cons of explainable AI.

Governance
Governance for AI touches fundamental values such as equality and fairness. Several participants called for the development of a vision of how society should work, including for AI and its use. However, a problem that was raised is that the interpretation of these values changes over time. What was morally the norm one or two generations ago is in many cases no longer valid. Humanity improves, but this makes it hard to create fixed rules that are built into systems that may exist over long periods. The interpretations are also complex, require delicate balancing of risks, and vary with different applications, domains, and communities where the same AI components are used.

Trust
Unlike other systems, the complexity of AI systems often makes them hard to grasp. AI systems are also usually trained in a way that is unusual in engineering, in that they minimise average error but provide no hard limits. (A bar of steel sold as 1 m plus or minus 1 mm will never be longer or shorter than that, whereas AI systems on average do not deviate more than a certain error, but may be wildly off in individual situations.) This is a huge obstacle in building trust. One way of handling this is to educate people. Another way of improving the situation is to educate AI ambassadors – people who can explain what a certain piece of equipment does and what to expect from it. They could also assist in procurement and in user studies. An issue that came up several times is expectation management. AI has often been grossly oversold; even the name AI is poorly chosen. In its defence it is hard to describe all the things that can be achieved using AI without it sounding like a universal solution for everything. A solution that was discussed heavily during the workshop is explainable AI, i.e. systems that can motivate their decisions when required. Such a property would be highly effective in building trust and setting the right expectations.

Skills
One conclusion that was almost unanimously agreed on during the workshop is the need for a common curriculum for AI education. This could include a minimum set of topics to be covered and practical exercises in AI applications going wrong. It could also be combined with a European certification as an AI engineer. Outside this, the workshop favoured more focus on the processes (see below) rather than focus on the individuals creating AI applications. It was also pointed out that ACM is currently working on standards for AI degrees.

Processes
There was a strong agreement in the workshop that the processes around the applications of AI are highly important. Good AI algorithms and well curated data are extremely important for a good AI application, and a good process surrounding these is critical. A principal purpose of a good process is to continuously improve quality. All agreed that there is vast experience that should be collected, curated and unified, compiled and condensed, and used for teaching and continuous improvement. However, there was some debate about how this should be done, since many perceive that there is a general reluctance in reporting, due to stigma and financial risk. There is also a large amount of work involved in curating and condensing the information. One proposal is to use the data factories (i.e. experimental facilities that facilitate the access to data for the development of methods and applications) to gather and disseminate best practice. A complementary proposal is an official European Commission publication, acting as an official channel to communicate best practices from multiple sources. Looking further into the future, one could also try to establish a research field in the use of AI, distinct from the development of AI, in order to stress the importance of this. Another suggestion was the creation of standard benchmarks. The difficulty here is maintaining an up-to-date set, and how strongly dependent on the application domains that the benchmarks will be. A solution would be to rely on institutes and universities to maintain this under the patronage of the European Commission. A point of strong agreement is that AI is software and that it is almost always a part of a large software system. In many respects we need to view the application of AI as a process and not as a product. We therefore need a software engineering approach to the use of AI, contrary to the mostly mathematical and algorithmic approaches used so far.

Testing
Testing and validation are integral parts of AI, and there is a large body of publications on this with respect to core AI. However, testing the entire systems that AI is part of is a different matter. This is especially difficult since the effects in the surrounding systems stemming from errors in the AI part are difficult to anticipate. It is also impossible to exhaus-

tively test the core AI module once it has even a moderate number of inputs.

While the "unknown unkowns" will always exist, their effects can be reduced with good processes. Some interesting questions evolved from the concept of the contract between the AI specialist and the problem owner. What are component's properties that the AI specialist delivers? How should they be described? What does it mean legally? How would one test for legal conformity against what is actually statistical properties in high dimensions? These are questions that need to be answered.

Quality of data and methods

The quality of data is an obvious weakness of AI. It is especially true if the system continuously learns from new data, as there is generally poor control over bias in this stream, and according to experience, the quality of data tends to deteriorate over time. Many of the participants advocated requiring AI providers and users to openly publish all their data. This was criticised based on issues of individual privacy and, business considerations. The proposed solution is to create AI auditors that will perform audits in a similar way to financial audits but on processes and data practices related to AI. It may be possible to create standards for training data with corresponding certification ("Only organic data used"). Given the rapid development in the area, it was suggested that this should be done by industry consortia or institutes, as formal standardisation processes would be too slow. If data is shared in several steps, a pedigree of the data needs to be established. A complementary approach that was suggested, is to provide good public datasets on which systems can be trained. A question is then how to maintain quality and relevance over time. This would require a curator, which could be – as mentioned above – under the auspices of the European Commission and might be part of the mission of data factories. There also need to be best practices on how to use data for training, testing and validation, the importance of cross-validation and permutation, etc.

Please note that the ideas presented do not necessarily reflect the opinions of any individual participant in the workshop.

Participants

The participants of the workshop were:
- Björn Levin, RISE
- Gabriel David, INESC TEC
- Daniel Gillblad, RISE
- Arnaud Gotlieb, Simula Research Laboratory
- Olivier Grisel, Inria
- Alípio Jorge, INESC TEC
- Bert Kappen, Radboud University
- Fabio Martinelli, CNR
- Michael Mock, Fraunhofer IAIS
- Anirban Mukhopadhyay, TU Darmstadt
- Ana Paiva, INESC-ID
- Han La Poutré, CWI
- Andreas Rauber, TU Wien

Please contact
Björn Levin, RISE, Sweden
ERCIM president
Bjorn.Levin@ri.se

25th International Conference on Formal Methods for Industrial Critical Systems

by Maurice ter Beek

The yearly conference of the ERCIM Working Group on Formal Methods for Industrial Critical Systems, FMICS, reached its 25th edition this year. A panel "Genesis, Success Stories, and the Next 25 Years" was organized to celebrate this remarkable milestone. The conference took place virtually on 2-3 September, hosted by TU Wien.

The aim of the FMICS conference series is to provide a forum for researchers interested in the development and application of formal methods in industry. The conference was chaired by Maurice ter Beek (ISTI-CNR, Italy) and Dejan Ničković (AIT, Austria) and organized under the umbrella of QONFEST 2020 organized by Ezio Bartocci (TU Wien, Austria). FMICS attracted a record number of 149 participants from many countries worldwide. QONFEST attracted over 400 participants. A good number of 26 papers were submitted, of which eleven were accepted. The program included three keynote lectures, partially shared with the QONFEST conferences CONCUR, FORMATS and QEST, by Roderick Bloem (TU Graz, Austria), Thomas Henzinger (IST, Austria), and Stefan Resch (Thales, Austria).

Following a tradition, Springer provided an award for the best FMICS paper. This year, the reviewers selected the contribution "Verifiable and Scalable Mission-Plan Synthesis for Multiple Autonomous Agents" by Rong Gu, Eduard Enoiu, Cristina Seceleanu, and Kristina Lundqvist for the FMICS 2020 Best Paper Award. The panel celebrating the 25th anniversary of FMICS was a big success.The founders and previous chairpersons of the ERCIM WG FMICS acted as panelists. Diego Latella (ISTI-CNR) recalled the original motivation and beginning of FMICS, Stefania Gnesi (ISTI-CNR) shared some success stories, and Hubert Garavel (Inria) presented a study on the future of formal methods and their adoption in industry. The detailed report [1] of this study, included in the proceedings [2], presents an analysis of the opinions of 130 renowned experts in formal methods, as well as thought-provoking position statements on formal methods of 111 of them.

Link: https://fmics20.ait.ac.at/

References:
[1] H. Garavel, M.H. ter Beek, and J. van de Pol: "The 2020 Expert Survey on Formal Methods", in [2], pp. 3-69. DOI: http://dx.doi.org/10.1007/978-3-030-58298-2_1
[2] M.H. ter Beek and Dejan Ničković (eds.): "Formal Methods for Industrial Critical Systems: Proc. of FMICS'20", Springer LNCS, 2020. DOI: http://dx.doi.org/10.1007/978-3-030-58298-2

Please contact:
Maurice ter Beek, ISTI-CNR, Italy,
maurice.terbeek@isti.cnr.it

Introduction to the Special Theme:

Blue Growth

by the guest editors Alberto Gotta (ISTI-CNR) and John Mardikis (EPLO - Circular Clima Institute)

Ocean-based industrial development and growth are important to the economic and social development of many countries, and both the EU and the US agree on the urgency of implementing an integrated maritime strategy with the aim of coordinating policies for the different sectors of the sea. The growth of the maritime economy (blue growth) is an opportunity that Europe cannot wait to seize [1] to create new job opportunities, support systemic competitiveness, and strengthen social cohesion. This approach is also fully aligned with the objectives of the UN Agenda 2030 for Sustainable Development, in particular, Objective 14 (SDG 14): "Conserve and sustainably use the oceans, seas and marine resources for sustainable development". International strategic documents [1] define the areas of blue growth as: coastal and maritime tourism, blue energy (renewable energy from offshore wind farms, onshore/offshore wave energy, salinity gradient, marine current, ocean thermal energy, biomass), marine environment and coastal strip preservation, abiotic and biotic marine resources, fisheries and blue biotechnologies, shipbuilding and marine robotics, smart ports, sustainable maritime transportation, maritime surveillance and safety, sustainability and economic uses of the sea.

The EU Blue Economy Report prepared by the Directorate-General Maritime Affairs and Fisheries (DG-MARE), provides "support to policymakers and stakeholders in the quest for a sustainable development of oceans, coastal resources and, most notably, to the development and implementation of policies and initiatives under the European Green Deal in line with the new approach for a sustainable Blue Economy".

Accordingly, the EU policies are designed to reinforce the efforts of the Member States and regions and provide common building blocks for a blue economy. EU funding under the 2014–2020 financial framework can reinforce these efforts. Ongoing EU initiatives are already encouraging innovation in different sectors and have identified the following five value chains that can deliver sustainable growth and jobs in the blue economy: blue energy, aquaculture, maritime, coastal and cruise tourism, marine mineral resources, and blue biotechnology [2].

Most of the EU funding programmes have the scope to finance projects in blue growth. The European Commission and the DG-MARE fund innovative projects and regularly announce calls for submissions. The European Maritime and Fisheries Fund (EMFF) is focused on seas and coasts, supporting the diversification and sustainable development of maritime economies. The European Regional Development Fund (ERDF) offers numerous opportunities for investing in blue growth, supporting programmes (such as INTERREG and Operational Programmes in each Member State) in the subsectors of sustainable tourism, protection of biodiversity, renewable energy, marine transport and fisheries. Many calls within the framework of "Horizon 2020", EU's largest research and innovation Programme, are focused on blue energy and new technologies in blue growth. Projects co-funded by the European Social Fund (ESF) could assist in strengthening training and education in the marine and maritime sectors, while the Cohesion Fund (CF) supports trans-European networks and environmental projects. The "Competitiveness of Enterprises and SMEs" (COSME) Programme provides financing opportunities to SMEs within blue growth. The "Connecting Europe Facility" (CEF) Programme finances blue growth projects related to port infrastructure, lowering the carbon footprint and compliance with air quality legislation. The EU Emissions Trading System (EU ETS) and the "NER 300" Programme support the deployment of innovative renewable energy technologies and low carbon energy demonstration projects.

While information and communication technologies are being increasingly widely used in industry and transportation, the maritime sector can also benefit from embracing the most promising technological solutions for the blue economy.

Open innovation model for the blue economy

An open innovation model (OIM) is a paradigm under which companies develop competency networks that link suppliers, universities and research centres to cooperate in solving complex problems. Campana and Ferro outline how the external knowledge of the CNR provided fuel to shipbuilding company Fincantieri's business model, enabling research and development to be converted into commercial value. Fincantieri applied a "coupled innovation process", a variant of the OIM, to six projects in the maritime field. Each project addresses a different technological issue within the maritime sector. The E-Cabin project creates a set of advanced technological solutions for cruise ship cabins to improve the onboard experience of passengers. The PiTER on-Board (Technological Platform for High-Efficiency Waste-to-Energy Thermo-Conversion on-board) project focuses on evaluating and developing poly-generation systems for onboard energy use and storage to deal with sustainable maritime transportation. The High-efficiency Vessel project is devoted to designing and conducting experiments on an advanced energy system, to increase the overall energy efficiency of ships, and the GEI Innovative Electric Generation project aims to develop new technologies to improve the efficiency and sustainability of the ship's electrical power plant, considering different aspects, including electrical generation, power distribution and energy management. The E-Navigation project addresses maritime surveillance and safety by developing a virtual dashboard showing digital information to support navigation and reusing a piece of it to suggest an adequate control of propulsion and navigation route, in order to avoid collisions with other objects sighted at sea. These features increase the number of operations that the ship can perform without the direct intervention of a human operator. The Secure Platform project addresses maritime surveillance and safety through two objectives: (i) To develop an advanced security system to protect passengers and personnel in both routine situations and emergencies. (ii) To realise a completely novel system to search and assist a passenger overboard.

Big data infrastructure

The BLUE-CLOUD flagship project of the Directorate-General (DG) for Research and Innovation Unit of the European Commission is establishing a thematic marine cloud serving the blue economy, marine environment and marine knowledge agendas and the European Open Science Cloud. The project links the horizontal e-infrastructures supported by DG-CONNECT and DG-GROW, long-term marine data initiatives supported by DG-MARE, research infrastructures supported by DG for Research and Innovation Unit and other recently funded thematic clouds. BLUE-CLOUD is also developing pilot demonstrator applications with the goal of establishing a marine-themed European Open Science Cloud (EOSC) for the blue economy and marine environment. "Fish, a matter of scales" is one of these demonstrators that aims to improve data management and analytical capabilities of fisheries. These topics are addressed in articles by Candela and Pagano as well as Tzitzikas and Marketakis.

Pendleton's paper presents the Centre for the Fourth Industrial Revolution – Ocean. The Centre, supported by the World Economic Forum, was established to pilot Fourth Industrial Revolution (4IR) solutions to problems facing the world's oceans. The Ocean Data Platform (ODP) is a new data infrastructure to pilot and scale data-oriented solutions to help chart a sustainable blue economy.

Planas-Bielsa et al. describe the World Coral Conservatory (WCC), which will function as a global network, sharing biological material via a big data platform based on a big data structure. It will contribute to the development of biomedical and cosmetic applications and will help to conserve coral reefs as an ecological and economic resource for future generations.

Toward intelligent autonomous and connected systems

Digital connection and situational awareness are becoming increasingly important in the naval field with the development of autonomous systems. The ship of the future will rely on a range of technologies, including satellite communications, navigation monitoring and control systems, sensor networks, machine-learning solutions, artificial intelligence techniques and prediction models. These technologies, which have the potential to improve the safety, security, efficiency, and sustainability of the maritime industry, are discussed in several articles in this issue.

Martelli et al. describe a distributed computing platform that enables automatic control for maritime services, with likely economic and social benefits. In this context, the nodes involved in the computing tasks are autonomous complex cyber-physical systems, i.e., ships. The platform allows node computing cooperation through a high-level abstraction of the underlying sensor system. The computing tasks are related to predictive analysis, employing artifi-

cial intelligence techniques based on the federated-learning paradigm.

New technology will also improve maritime safety. Mpeis et al. give an overview of A4IoT, an innovative localisation architecture that supports a smart alert system to provide monitoring, navigation and guidance to first responders during fire outbreaks on roll-on/roll-off vessels.

The project "Intelligent ICT Applications for the Management of Marinas and Yachts" is using intelligent approaches to address the unique challenge of providing journey-planning tools to yachters. Yachting tourism has the potential to drive strong blue growth in hosting countries due to its multiplicative effects in other related service and goods economy sectors (Komninos et al.).

While technology will have many benefits for the blue economy, it will also expose maritime systems to new security and safety risks in the form of cyber-attacks. Attacks on maritime ports have become more sophisticated since modern ports turned into cyber-physical systems. König describes a simulation model that can help with the vital task of detecting and understanding the impacts of such attacks. The model, developed within the European Commission's project SAURON, simulates the aftermath of a security incident in a port in order to understand the impact on the port and the local population.

The design of efficient seagoing vessels is vital for the growth of a sustainable ocean economy. Ship performance analyses become more meaningful if they report the uncertainty quantification analysis, which means that several scenarios should be tested, and results condensed in statistical terms. Recently developed multi-fidelity, adaptive algorithms can provide these figures promptly, and further research in this area is needed to make these methods even more reliable, effective and ready for the general public (Broglia et al.).

Marine heritage monitoring

Some projects are focusing on collecting as much information as possible about marine resources, with the aim of helping researchers and industry to develop new environmentally sustainable products for human use.

The NAUTILOS project aims to fill some of the existing gaps in marine observation and modelling by improving the measurement of chemical, biological, and deep ocean physics variables. Pieri explains how these will complement and expand existing observation tools and services, allowing researchers to obtain data at a much higher spatial resolution, temporal regularity and length than what is currently available.

Remote sensing provides almost global spatial coverage, but with limits in resolution and accuracy. Ground stations, conversely, provide very accurate coverage with a high temporal resolution, but sparse and point-wise. What's missing from this picture is accurate local knowledge with a high spatial resolution, making it possible to understand and analyse local phenomena to their full spatial extent. To fill the gap, Berretta et al. propose a paradigm shift in field sampling, environmental surveys that are dramatically cheaper, quicker and easier to perform, and the ability to perform visualisation and analysis tasks as the survey progresses. This is made possible by a real-time adaptive sampling method, embedded within a mobile observation platform.

The global demand for hydrocarbons and other mineral resources is increasingly met by tapping the vast seabed resources in ever more difficult and risky environments, like the abyssal oceanic depths and the Arctic. Andritsos discusses how technology can help improve the safety of underwater operations to exploit these resources in a sustainable manner. This is particularly important in the delicate polar environments and the environmentally and socially stressed Eastern Mediterranean regions.

This ERCIM News special theme is devoted to the application of ICT technologies to the blue growth field, which includes (but is not limited to) machine learning and artificial intelligence, cyber-physical systems of systems, IoT/M2M communications, space communications and observations, big data infrastructures, unmanned and autonomous systems.

References:

[1] National Technological Cluster Blue Italian Growth - CTN BIG - Action Plan 2014-2020

[2] European Commission, Blue Growth – Opportunities for marine and maritime sustainable growth – Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the regions, Luxembourg: Publications Office of the European Union 2012—13pp. ISBN 978-92-79-25529-8 DOI:10.2771/43949

Please contact:

Alberto Gotta
ISTI-CNR, Italy
alberto.gotta@isti.cnr.it

John Mardikis
EPLO - Circular Clima Institute, Greece
jmardikis@circularclima.org

An Open Innovation Model: Six Projects in the Maritime Field

by Emilio Fortunato Campana (DIITET-CNR) and Erina Ferro (ISTI-CNR)

One of the largest shipbuilders in the world, Fincantieri (FC), together with Italy's largest research institution, the National Research Council (CNR), successfully applied an open innovation model (OIM) to six projects in the maritime sector. The external knowledge of the CNR provided fuel to Fincantieri's business model, enabling research and development to be converted into commercial value.

An open innovation model (OIM) is a paradigm under which companies develop competency networks that link suppliers, universities, and research centres to cooperate in solving complex problems. The shipbuilding company Fincantieri (FC) together with researchers from Italy's National Research Council (CNR) have adopted a "coupled innovation process", a variant of the OIM [1], simultaneously applying it to six different projects. Each project addresses a different technological issue within the maritime sector.

The CNR coordinated the six projects outlined below, acting as an innovation "hub". CNR was the interface between the public research system and industry, thus being responsible for selecting the skills, collecting the results and, with the support of FC, collating summaries of results into an overall road-map. The work involved nine institutes of the CNR-DIITET department (Department of Engineering, ICT and Technologies for Energy and Transportation), one institute of the CNR-DSCTM department (Department of Chemical Science and Materials Technologies) and three universities (Genoa, Trieste and Rome-La Sapienza).

E-Cabin (Figure 1) created a set of advanced technological solutions for cruise ship cabins to improve the onboard experience of passengers. The solutions are personalised and include: (i) A cabin monitoring system comprising a set of heterogeneous sensors operating inside the cabin. The sensors evaluate energy consumption, schedule cabin maintenance, measure environmental variables, and automatically actuate solutions appropriate for the specific passenger in that cabin; (ii) A set of applications that learn an individual passenger's habits. These applications predict the individual's needs, maximising each passenger's opportunities to socialise by sharing information through mobile social networking applications. Augmented reality contents relevant to the cruise also help passengers to participate in the "ship world"; (iii) A set of applications based on augmented reality to facilitate passengers' movements within the ship.

A study on the comfort perceived by the passenger, when multiple disturbing factors act simultaneously, is also included. We also developed an energy harvesting solution to power sensors. A dashboard is provided to the ship's technical staff to visualise the status of each cabin and receive information about all the monitored parameters.

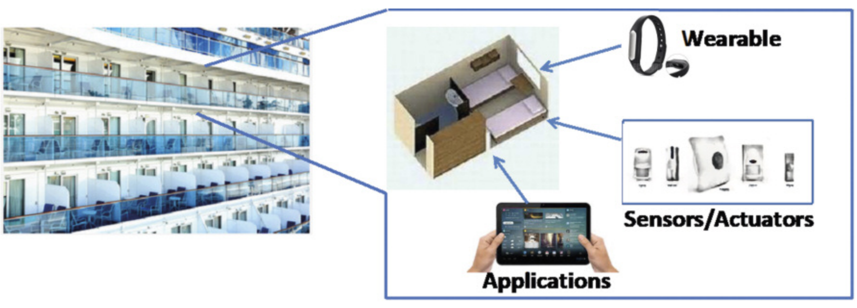


Figure 1: The E-Cabin system.

PiTER on Board (Technological Platform for High-Efficiency Waste-to-Energy Thermo-Conversion on board) focused on evaluating and developing poly-generation systems for onboard energy production and storage. This system derives its primary energy supply for a bottoming thermodynamic cycle from residual biomass and/or engine waste heat produced during the ship's normal operation. The organic fractions of food waste, waste vegetable oils, and wastewater (sludge) are taken into consideration to produce biogas, syngas, and liquid fuels to use on board. Energy management strategies and CO_2 capture are considered, to achieve optimal integration of each component into the onboard energy grid. Different energy conversion systems and schemes have been evaluated to identify the most suitable for the complete waste-to-energy system. The aim was to optimise the performance of the system in terms of fuel flexibility, global efficiency, low environmental impact, and suitability for waste heat recovery and CO_2 capture.

High-efficiency Vessel is designing and conducting experiments on an advanced energy system to increase the overall energy efficiency of ships by recovering waste heat from the ship's propulsion system. The potential of combining organic rankine cycle (ORC), Stirling engines and latent thermal energy storage systems have been assessed, and a lab-scale prototype of the energy recovery system has been constructed. The propulsion system of the ship was simulated with a small diesel engine. Other core components are a thermal storage tank prototype with embedded phase change materials, a commercial ORC system and a Stirling engine, optimised for the specific application. The exhaust line of the diesel engine was endowed with a proper design gas-to-gas heat exchanger to provide thermal energy to the Stirling engine. Moreover, the cooling system of the diesel engine was modified to send the hot water from the engine alternately to the ORC or to the thermal storage, to maximise heat recovery. An advanced control tool was developed to manage and optimise the

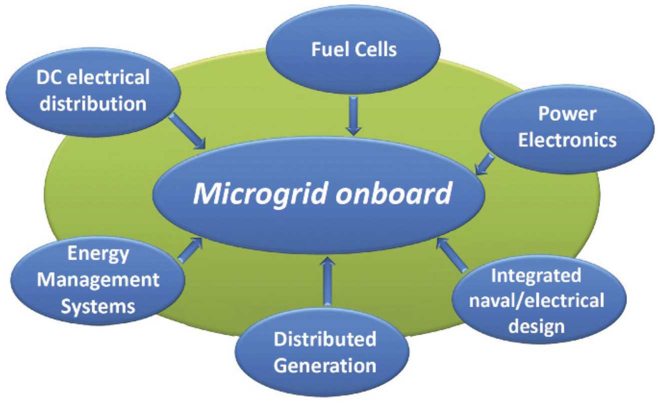


Figure 2: The main outcomes of the GEI project.

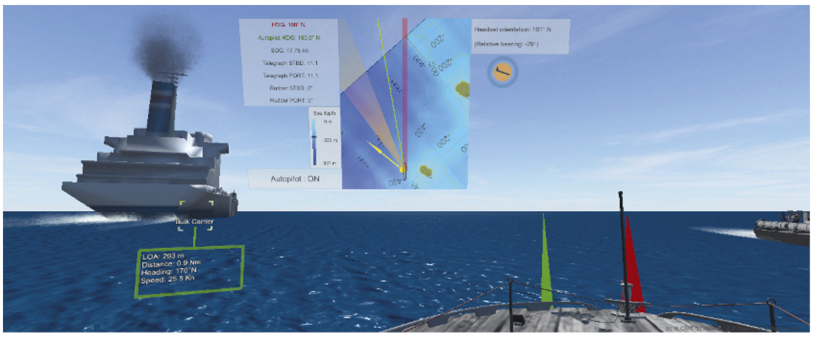


Figure 3: The E-Navigation project.

energy flows as a function of user demand.

GEI (Innovative Electric Generation) (Figure 2) aimed to develop new technologies to improve the efficiency and sustainability of the ship's electrical power plant, considering different aspects including: electrical generation, power distribution and energy management. How to maximise the payload and increase the system safety and reliability were also investigated. Five main outcomes were: (i) The design of new direct current (DC) and hybrid medium voltage alternating current (MVAC)/medium voltage DC (MVDC) electrical architectures; (ii) The design of a new electrical generation system based on fuel cells; (iii) The definition of a new paradigm of distributed generation on board; (iv) The development of new energy management techniques to optimise the use of electrical power on board; and (v) The development of new, reliable, efficient, and compact power electronic converters.

E-Navigation (Figure 3) developed a virtual dashboard for using digital information to support navigation, as well as some features of the propulsion control and navigation. These features increase the number of operations that the ship can perform without the direct intervention of a ship operator. The augmented reality system allowed the digital overlapping of information over the perceived reality. The displayed information comprised routes, speed, possible obstacles on the route and their characteristics, captions of naval vehicles near the target ship, and information on control commands of the propulsion. The augmented information is used via smartphones, tablets, smart glasses or binoculars.

Secure Platform had two main objectives: (i) To develop an advanced security system to protect passengers and personnel in both routine situations and emergencies. This was achieved with computer vision techniques based on cameras (visible and thermal) for people and goods tracking. We also used multisensory biometric recognition techniques (fingerprint, vocal imprint, face recognition) to restrict access to certain areas on the ship. In addition, radar technologies were used to detect, localise and track people within a closed environment; (ii) To realise a completely novel system for search and assistance of a passenger overboard. This system was based on: a) air drones to raise the alarm and to initialise the search if a passenger falls overboard. If the individual is uncooperative, one or more drones can be automatically launched to locate the person and activate the recovery operation; b) autonomous marine unmanned robotic vehicles, integrated with the man overboard recognition and tracking subsystem, for the rescue of the person.

All these projects started in January 2017 and ended in January 2019. During the collaboration between FC and CNR, new services were developed in each of the six projects, confirming the importance of interdisciplinary collaboration for the successful development of new innovative services. Innovation in services increases the quality, contributes the skills of new services, answers to the requests, creates new services, and can increase the competitive advantage because it contains the knowledge of customers and employees. It also implements new markets for services and it promotes competitive differentiation between competitors. The model used is a perfect example of harmonious balance between the inbound and outbound open innovation process. The ideas from outside (CNR and universities) were conveyed inside Fincantieri and combined with their internal ideas to generate innovative outcomes, for the benefit of both Fincantieri and the research units. Moreover, all resources necessary to successfully complete the program of the six projects were available, as, in total, there were over sixty researchers involved, thus guaranteeing the alignment between the necessary resources and the program to carry on. The model applied was so successful that Fincantieri is replicating it in another innovation project devoted to develop technologies that can prove to be disruptive for the maritime industry.

Reference:

[1] E. F. Campana, et al.: "CNR–Fincantieri Joint Projects: A Successful Example of Collaboration between Research and Industry Based on the Open Innovation Approach", Journal of Open Innovation: Technology, Market, and Complexit, vol. 6, issue 1, 2020.

Please contact:

Erina Ferro
ISTI-CNR, Italy
erina.ferro@isti.cnr.it

Understanding and Managing Ocean Sustainability: The Blue-Cloud Project

by Leonardo Candela and Pasquale Pagano (ISTI-CNR)

The Blue-Cloud flagship project of the Directorate-General (DG) for Research and Innovation Unit of the European Commission is establishing a thematic marine cloud serving the blue economy, marine environment and marine knowledge agendas and the European Open Science Cloud. The project links the horizontal e-infrastructures supported by DG CONNECT and DG GROW, long-term marine data initiatives supported by DG MARE, research infrastructures supported by DG for Research and Innovation and other recently funded thematic clouds.

The oceans, seas, and coasts are home to diverse marine ecosystems, provide a wealth of resources, influence climate, and offer many economic opportunities. However, marine ecosystems are sensitive to long-term global change and combinations of local pressures can affect ecosystems in unpredictable ways. A thorough understanding of the natural dynamics and effects is needed in order to develop sustainable management strategies and conservation of these ecosystems. To support the required research, we need a platform that offers seamless access to marine data as well as computing and required services. The Blue-Cloud project [L1] is addressing this goal through a set of five compelling demonstrators that address societal challenges:

- *Zooplankton and Phytoplankton Essential Ocean Variables Products*, led by the Marine Flanders Institute (VLIZ), is building upon a range of oceanographic data from multiple streams to produce and share unique 3D and 4D synergistic zooplankton and phytoplankton products. The products will contribute knowledge (they represent a class of ocean variables to be measured considered by the scientific community as crucial for ocean observation in a global context) helping to quantitatively reduce uncertainty about the present state of the marine plankton ecosystems and their response to climate change.
- *Plankton Genomics*, led by the European Molecular Biology Laboratory (EMBL), is building an environment enacting a deep assessment of plankton distribution, dynamics and fine-grained diversity to molecular resolution, focusing on discovery of species and functions and exploring genetic and morphological markers of plankton diversity and abundance.
- *Marine Environmental Indicators*, led by the Euro-Mediterranean Center on Climate Change (CMCC), is developing an online service with associated cloud based analytical computing framework and dedicated web interface to provide and display indicators and information on the environmental quality of the ocean.
- *Fish, a matter of scales*, coordinated by the Food and Agriculture Organisation of the UN (FAO), is improving data management and analytic capabilities for fisheries by building a global vertically integrated toolset to manage public fisheries' statistical data from ingestion, and harmonisation through to publication.
- *Aquaculture Monitor*, coordinated by the Food and Agriculture Organisation of the UN (FAO), is using Copernicus data, and combining AI with in-situ datasets to obtain regional-level inventories of aquaculture activities. This will enable the development of a robust and replicable environment for monitoring aquaculture in marine cages and coastal areas.

Figure 1: Blue-Cloud Federation of European Infrastructures.

The Blue-Cloud platform is neither built from scratch nor operating in a vacuum; it is designed to leverage existing and forthcoming data sources, infrastructures, and services operated by diverse providers under different settings. It will offer a unified, yet evolving, working environment that provides access to and use of the aggregated assets. Moreover, it uses and complements the resources of the European Open Science Cloud (EOSC) [L2], i.e. the platform the European Commission is supporting to help scientists find and access data and services across a plethora of providers. The Blue-Cloud platform will contribute data and services to expand on EOSC's existing system.

The Blue-Cloud platform aggregates assets from major domain specific data providers (Figure 1), including SeaDataNet (marine environment), EMODnet Bathymetry (bathymetry), EMODnet Chemistry (chemistry), EurOBIS –EMODnet Biology (marine biodiversity), Euro-Argo and Argo GDAC (ocean physics and marine biogeochemistry), ELIXIR-ENA (biogenomics), EuroBioImaging (microscopy), EcoTaxa (bio images), WekEO (CMEMS ocean analysis and forecasting and C3S climate analysis and forecasting), and ICOS-Marine (carbon). In addition to this content-centric approach, the platform offers a rich array of facilities to support the entire lifecycle of a research workflow with dedicated working environments built on the D4Science infrastructure [1].

D4Science, which has been operating since 2006, is an infrastructure that facilitates the development and operation of virtual research environments (VREs) for several "communities of practice", from domains including agri-food, earth science, marine science, social sciences, and humanities. Each VRE is a dedicated working environment designed to serve the needs of its community. D4Science-based VREs are web-based, community-oriented, collaborative, user-friendly, open-science-enabler working environments for scientists and practitioners who are collaborating on research tasks.

From an end-user perspective, each VRE manifests as a unifying web application (and a set of application programming interfaces (APIs)) that: (i) comprises several components and (ii) runs in a plain web browser. Each VRE component transparently provides the user with services and facilities possibly operated from diverse providers. In fact, every VRE acts as a gateway that provides seamless access to relevant datasets and services while hiding their diverse origins. Each VRE offers some basic components that help users collaborate [1], namely: (i) a workspace to organise and share digital artefacts; (ii) social networking to communicate with co-workers using posts and replies; (iii) data analytics to share and execute analytics methods using a transparent distributed and heterogeneous computing infrastructure; (iv) a catalogue to document and publish any digital artefacts. To help users integrate their assets into the VRE, three integration patterns are supported (besides implementing completely new services) [2]: (i) integration of existing applications; (ii) integration of analytics methods and workflows; and (iii) integration of datasets and other resources for discovery and access.

The five demonstrators are building specific working environments, services, tools, and datasets and making all of this available by a dedicated gateway [L3].

In addition to technical work, the project is developing a pragmatic, policy-oriented roadmap to 2030, undertaking extensive stakeholder consultations with the wider marine and ICT communities. The goal is to co-design the future strategic development of the wealth and diverse infrastructures and initiatives for collecting, managing and providing access to marine data.

In summary, Blue-Cloud showcases the development of a platform for marine scientists and stakeholders by intertwining a set of interoperable infrastructures and provides insights on how to further develop the resulting platform in the future.

Links:
[L1]: https://www.blue-cloud.org/
[L2]: https://kwz.me/h1Z
[L3]: https://blue-cloud.d4science.org/

References:

[1] M. Assante et al.: "Enacting open science by D4Science", Future Generation Computer Systems, Volume 101, 2019, Pages 555-563, https://doi.org/10.1016/j.future.2019.05.063
[2] M. Assante et al.: "Co-creation of Virtual Research Environments: the D4Science Experience". 12th International Workshop on Science Gateway (IWSG2020), 2020

Please contact:

Pasquale Pagano
ISTI-CNR, Italy
pasquale.pagano@isti.cnr.it

Reinforcing Fisheries Management through Semantic Data Integration

by Yannis Tzitzikas and Yannis Marketakis (FORTH-ICS)

The European project BlueCloud is developing pilot demonstrator applications with the goal of establishing a marine-themed European Open Science Cloud (EOSC) for the blue economy and marine environment. "Fish, a matter of scales" is one of these demonstrators that aims to improve data management and analytical capabilities of fisheries.

Fisheries management is a laborious task that relies on data analysis using complex models and fine-grained software over several sources of information in order to deduce certain facts with the overall aim of improving the sustainability of fisheries. It includes the usually manual process of identifying and combining different parts of information, which is an extremely time-consuming and error-prone process. The key indicators for efficient fisheries management are stocks and fisheries. Stocks refer to groups or individuals of a species occupying a well-defined spatial range (e.g. swordfish in the Mediterranean Sea). Fisheries describe the activities leading to the harvesting of the fish within a particular area, using a particular method or equipment and purpose of activity (e.g. the Atlantic cod fishery in the area of East and South Greenland). The knowledge of the status and the trends of stocks and fisheries at regional, national and local levels is the key factor for highly reliable fisheries management.

To this end, the Global Record of Stocks and Fisheries (GRSF) [1], developed within the context of the EU H2020 project BlueBRIDGE (GA no: 675680, 2015-2018), has collated stocks and fisheries information from three distinct data sources: FIRMS from the Food and Agriculture Organization of the United Nations (FAO), RAM Legacy Stock Assessment database and FishSource from the Sustainable Fisheries Partnership. These sources were chosen because they contain complementary information both conceptually and geographically. By collating these sources, the reporting coverage of any of the single entities is increased. To achieve

Figure 1: (upper) GRSF semantically integrates data from heterogeneous sources and can be expanded with more data sources through the BlueCloud Discovery and Access facilities. (lower) GRSF information can be discovered and exposed through the catalogues of a dedicated Virtual Reseach Environment within BlueCloud project.

this, we have defined a workflow of activities that semantically integrate these sources in order to deliver a single source of information (GRSF).

The original data sources use different data models and formats to store and expose their information, as well as different terminologies and standards. Underpinning the GRSF is a set of standards and rules agreed upon by the stakeholders of the database sources. For example, stakeholders have agreed on the use of FAO ASFIS (3-alpha) codes for specifying species (i.e. SWO is the 3-alpha code for swordfish), and that the species involved and the occupying water area are the unique fields that define a single stock. The reliance on standards and the fields that define the uniqueness of records have a direct effect on the generated semantic identifiers of records. These identifiers are formulated in a way that can be understood by both humans and computers. Each identifier is a concatenation of a set of predefined fields of the record in a particular form. To keep them as short as possible it has been decided to rely on standard values or abbreviations whenever applicable. Each abbreviation is accompanied with the thesaurus or standard scheme that defines it. For example, the semantic identifier "ASFIS:SWO+FAO:34" is used to identify a stock record about the species with code SWO with respect to FAO ASFIS standard (i.e. its common name in English is swordfish), in the area with code 34 with respect to FAO fishing areas coding scheme (i.e. this area is known as the eastern part of the Atlantic Ocean). On the contrary, for fishery records more fields are used to generate the semantic ID, and consequently the uniqueness of a fishery record, namely: the species, the water area, the management authority, the fishing gear used and the country under which the fishery is operated (e.g. ASFIS:COD+FAO:21.3.M+authority:INT:NAFO+ISSCFG:03.1.2+ISO3:CAN).

The proposed workflow includes concrete steps for harvesting data from the remote data sources, normalising them both syntactically and semantically, applying schema mappings between the schemata of the different data sources and MarineTLO ontology [L3], transforming them to ontological instances of MarineTLO, applying merging and dissection rules in order to deliver a concrete set of stocks and fisheries records, and publishing them in the catalogue of a virtual research environment (VRE) [L2], operated within BlueCloud infrastructure, so that experts in fisheries management can assess them. Since the purpose of GRSF is not to substitute the underlying sources, they will continue to evolve independently. In order to harmonise their "fresh" contents with GRSF, we have also designed a refresh workflow that carries out all the aforementioned activities and also preserves all the manual edits and annotations made by GRSF administrators (e.g. proposals for merging records from different sources into a single GRSF record), as well as their public URLs assigned by the catalogue VRE.

As a global reference for the status and trends of stocks and fisheries, GRSF can help stakeholders involved in fisheries management achieve a better, more comprehensive and up-to-date view that will facilitate their decision making activities. For example, the FAO will be supported with the provision of certified traceability schemes for seafood products. In addition, other organisations, the industry and IT companies can build on top of GRSF to develop seafood traceability solutions based on standardised fishery identifiers.

As a follow-up, GRSF continues its expansion in the context of the existing H2020 BlueCloud project [L1] (GA no: 862409, 2019-2022) with information about the status assessment of fisheries, as well as with fish food and nutrition information. Such information will be derived from data sources developed in other projects and will be accessed through the data discovery and access services developed in BlueCloud project, as shown in the upper part of Figure 1. Overall, GRSF showcases a very promising domain-specific real-world application of processes for achieving large scale semantic integration [2].

Links:
[L1] https://www.blue-cloud.org/
[L2] https://kwz.me/h1G
[L3] https://kwz.me/h2o

References:
[1] Y. Tzitzikas et al: "Methods and Tools for Supporting the Integration of Stocks and Fisheries", in Int. Conf. on Information and Communication Technologies in Agriculture, Food & Environment, pp. 20-34, Springer, 2017.
[2] M. Mountantonakis and Y. Tzitzikas: "Large Scale Semantic Integration of Linked Data: A Survey", ACM Computing Surveys, 52(5), 2019.

Please contact:
Yannis Tzitzikas, FORTH-ICS and University of Crete, Greece
+30 2810 391621, tzitzik@ics.forth.gr

Yannis Marketakis,
FORTH-ICS, Greece
+30 2810 391629, marketak@ics.forth.gr

A New Data Infrastructure to Pilot Ocean Data Solutions

by Linwood Pendleton (Centre for the Fourth Industrial Revolution - Ocean)

The Centre for the Fourth Industrial Revolution – Ocean, an affiliate of the World Economic Forum, is partnering with the Ocean Data Platform to pilot Fourth Industrial Revolution (4IR) solutions to ocean problems. The Ocean Data Platform (ODP) is a new data infrastructure to pilot and scale data-oriented solutions to help chart a sustainable blue economy.

In the last 12 months, the Ocean Data Platform (ODP) has emerged as a new data infrastructure for piloting solutions to some of the key data sharing and access problems that must be overcome to chart a course for a sustainable blue economy.

The Ocean Data Platform [L1] was originally conceived of and created by REV Ocean [L2], a not-for-profit company that is building the world's largest luxury research yacht of the same name. REV Ocean is dedicated to creating new science and technology to improve the health of the ocean and is set to become one of the world's most sophisticated and prolific producers of ocean data and knowledge. The ODP and REV Ocean work closely together, sharing knowledge, expertise, and sometimes human resources.

The ODP was created in part to deal with the huge variety and volume of data that will be produced by the research vessel, but the platform was also conceived with the understanding that the data challenges to be experienced by the REV Ocean are just a microcosm of those facing the use of ocean data worldwide. The ODP plans to scale solutions developed working with REV Ocean to a more global set of data providers and users. More recently, the mandate of the ODP has grown thanks to its merger with the Centre for the Fourth Industrial Revolution (C4IR) Ocean – an affiliate centre of the World Economic Forum's C4IR Network. The C4IR Ocean brings together partners from businesses, governments, academia, and conservation organisations to identify and solve challenges that might impede the application of 4IR technologies to ocean problems.

Why do we need to pilot 4IR solutions for the ocean?

The ocean is changing rapidly due to climate change and human pressure. Decisionmakers and researchers need more and better data to understand, monitor, and model ocean conditions.

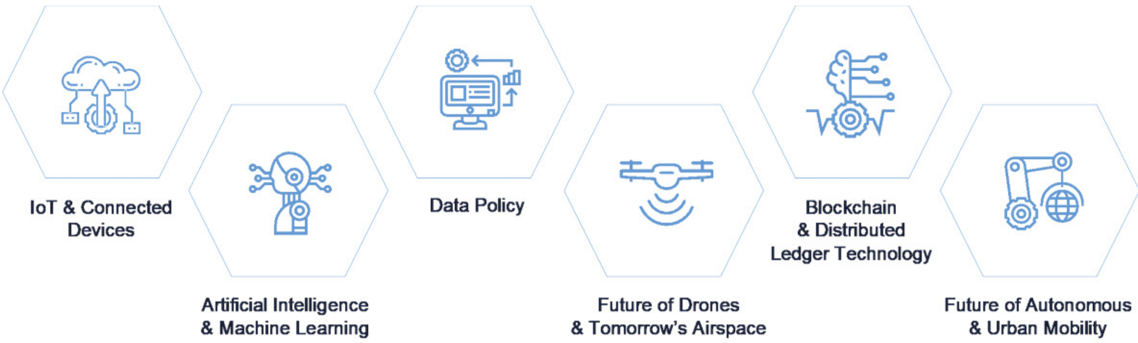


Figure 1: Centre for the Fourth Industrial Revolution focal areas.

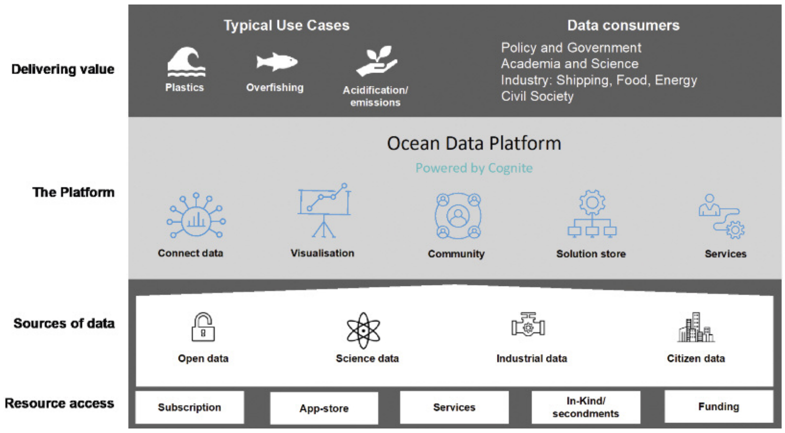


Figure 2: The Ocean Data Platform - data flow and infrastructure.

Ocean data are key to planning for sustainable development [L3], but also for industry to use the ocean and its resources and even to protect national security [1]. Yet, many aspects of the ocean remain poorly measured and monitored.

New fourth industrial revolution technologies have dramatically increased our ability to observe the ocean and the people who use it. Satellites can collect ocean data without regard to national borders and sovereignty. Autonomous devices can glean ocean data from below the surface, sometimes surreptitiously. Massive amounts of big ocean data from video, imagery, sound, radar, and a variety of new sources of data are emerging.

As the technology for collecting ocean data has rapidly expanded, so too has the number of types, formats, and standards being applied to ocean data. Getting data from source to user remains a huge challenge. Even when technological solutions exist for the liberation and sharing of data, much digital information – including open data – remains trapped in silos, prevented from flowing by cultural, proprietary, legal, and diplomatic issues. The goal of the ODP is to make more data open and to make open data easier to find and use.

Recognising the role of data for securing a healthy, safe, and productive ocean and the need to quickly break down these barriers, the United Nations called for a decade [L4] dedicated, in part, to a global transformation in how ocean data are collected and shared around the world. The C4IR Ocean was created, in part, to meet that call to action.

How does it work?

All C4IR affiliates work with the C4IR headquarters in San Francisco, California, to find governance solutions to challenges associated with six key areas of 4IR technology (See Figure 1). Governance solutions include voluntary approaches, best practices, agreed upon standards, and even regulation that may improve societal outcomes by improving the discovery and flow of data, interoperability and transparency, while also protecting privacy and intellectual property. All six areas of focus have direct application to the blue economy and sustainable ocean challenges.

The process followed by the C4IR Ocean has four key steps: (i) scan for challenges and opportunities, (ii) convene partners to brainstorm solutions, (iii) pilot solutions, and (iv) scale up. The Ocean Data Platform serves as an important tool for piloting data-oriented solutions.

The ODP is being built intentionally to meet the needs of those who demand more and better ocean data – specifically those app and tool builders, scientists, and modellers creating new sustainability solutions for planners, businesses, and others in the ocean sector. That means we need to build fast and efficient means of ingesting data, develop new ways of tracking data provenance, guarantee data integrity, and design better visualisation, navigation and data extraction tools. To do this, the ODP is built on Cognite Data Fusion which has been developed to handle massive amounts of industrial data, especially for maritime industries.

The data moving through the ODP will be drawn from a variety of sources including traditional, open, institutional ocean data as well as data from industry, independent scientists, and even private citizens (Figure 2). Trusted sources include (but are not limited to) data like those hosted by International Oceanographic Information and Data Exchange, the World Ocean Database, the Global Ocean Observing System, SeaDataNet, EMODNET, the EU's Copernicus and the USA's National Oceanic and Atmospheric Administration, and more specialised data like AIS and VMS vessel data, satellite imagery, acoustic, and other types of data. As we learn by working to pilot these C4IR Ocean use cases, we will continually add new data sources to the ODP.

Of course, the ODP is not alone in recognising the need for new ways to access ocean data. Many ocean data portals are being built around unique data sources or for specific purposes. As a true data platform, the ODP will integrate the good work of these portals and data sources and allow for the open source creation of new tools and analysis based on these combined efforts.

Links:
[L1] https://www.oceandata.earth/
[L2] https://www.revocean.org/
[L3] https://kwz.me/h1D
[L4] https://www.oceandecade.org/

Reference:

[1] L. Pendleton, K. Evans, M. Visbeck: "Opinion: We need a global movement to transform ocean science for a better world", proc. of the National Academy of Sciences 117 (18), 9652-9655, 2020
[2] J., M. Abbott, H. Sakaguchi et al.: "Technology, Data and New Models for Sustainably Managing Ocean Resources. Washing, DC: World Resources Institute", 2020. https://kwz.me/h1F

Please contact:

Linwood Pendleton
Centre for the 4th Industrial Revolution – Ocean, Norway
linwood.pendleton@oceandata.earth

The World Coral Conservatory: Offering New Hope for Corals

by Víctor Planas-Bielsa (Centre Scientifique de Monaco and Université Côte d'Azur), Rachid Benchaouir (Coraliotech, Monaco), Denis Allemand (Centre Scientifique de Monaco and Université Côte d'Azur) and Didier Zoccola (Centre Scientifique de Monaco and Université Côte d'Azur)

Climate change affects coral reefs worldwide. To counter the anticipated disappearance of coral reefs by 2100, the World Coral Conservatory was created as a "Noah's Ark" for corals. In addition to providing a living coral repository to protect and restore degraded reefs, the project will offer a global big data structure to supply resources for fundamental research into corals and to assist biotechnology companies in developing promising molecules for human health applications.

Reef-building corals are the animals that constitute the foundation of coral reef ecosystems that provide habitat and trophic support for one-third of all marine organisms. The World Coral Conservatory (WCC) is an international community of research laboratories, public aquaria, stakeholders (such as park authorities, local decision-makers, policy makers, NGO representatives), and national administrations that builds on the idea of using aquaria as a "Noah's Ark" to preserve corals. The Conservatory will function as global network, sharing biological material via a big data platform [1].

This collaboration is urgently needed, with climate change endangering many coral reefs (Figure 1), causing them to decline in size and biodiversity and with many species expected to face extinction by the end of the century. In addition to their ecological importance, coral reefs are economically essential for many human societies; more than 600 million people around the world depending directly on reefs for their survival. Moreover, corals possess numerous active substances that defend them against physical, chemical or biological threats. These substances could have promising applications, especially in the areas of human health and well-being [2].

Despite its central role in the regulation of metabolic activities, structural functions, homeostasis and defence, the cell proteome, i.e. the protein/peptide content of the body's cells at a given time, has not been used in many applications for human health or other areas. The loss of coral reef biodiversity could mean lost opportunities to discover and take advantage of many bioactive substances.

For this reason, it is critical that in addition to the living coral repository, the genetic information is also preserved using coral genomic and transcriptomic databases. These should provide an invaluable source of in-silico raw material that can be exploited at any time in the future to discover and valorise innovative polypeptide-based products. Anti-cancer molecules, antibiotics, sun protection and new anti-inflammatory agents are some examples of products that can feed the pipeline of biotechnology companies looking for new pharmaceutical molecules or cosmetic active ingredients.

Computing technology will be instrumental in finding these substances, with artificial intelligence (AI) showing great potential as a tool for drug discovery [3]. Advances in deep learning and computer power will very likely open the door for numerous applications to analyse and discover promising molecules based primarily on data mining of the genomic information. Such techniques are not meant to replace already-known pharmacological protocols, but to complement them in order to accelerate the rate of discovery and reduce the cost and risk. The approach requires the collaboration of: (i) research teams specialising in fundamental science like mathematics, biology, ecology and bioinformatics, and (ii) biotechnology companies that can ensure the transition from

Figure 1: Coral reefs are threatened by climate change (credit D. Zoccola/ CSM).

research and development to industrial production.

Fundamental science is important, particularly at this stage; while the AI approach is promising, much research is still needed to develop statistical and mathematical techniques to efficiently handle the enormous size and high dimensionality of the genomic datasets.

The role of biotechnology companies is equally vital. It is their job to guarantee the toxicological safety and efficacy of the selected molecules, to secure innovations through patent filings and to commercialise their solutions through licensing or selling innovative products.

In conclusion, the WCC, based on a big data structure, will contribute to the development of biomedical and cosmetic applications and will help to maintain coral reefs as an ecological and economic resource for future generations.

References:

[1] D. Zoccola et al.: "The World Coral Conservatory (WCC): a Noah's ark for corals to support survival of reef ecosystems. Plos Biology 2020, 18(9): e3000823. doi.org/10.1371/journal.pbio.3000823

[2] Sang VT et al.: "Coral and Coral-Associated Microorganisms: A Prolific Source of Potential Bioactive Natural Products. Mar Drugs. 2019;17(8):468, 2019. doi:10.3390/md17080468

[3] J. M. Stokes et al.: "A Deep Learning Approach to Antibiotic Discovery", Cell, Vol. 181, Issue 2, 2020, pp 475-483. doi.org/10.1016/j.cell.2020.01.021

Please contact:

Didier Zoccola
Centre Scientifique de Monaco, Principality of Monaco
zoccola@centrescientifique.mc

The Internet of Ships

by Michele Martelli (Università degli Studi di Genova), Pietro Cassarà (ISTI-CNR), Antonio Virdis and Nicola Tonellotto (UNIPI)

A distributed computing platform can provide automatic control for maritime services, with likely economic and social benefits. In this context, the nodes involved in the computing tasks are autonomous complex cyber-physical systems, i.e., ships. The platform allows node computing cooperation through a high-level abstraction of the underlying sensor system. The computing tasks are related to the predictive analysis, employing artificial intelligence (AI) techniques based on the federated-learning paradigm.

The best way to increase the efficiency and reliability of maritime services is to learn how to react automatically to what happens in the surrounding environment. Consequently, the platform must be able to tune the computation tasks parameters accordingly to the state in which the system will be. A novel approach to automation is to use a distributed platform that implements a digital characterisation of cyber-physical systems – in this case, unmanned surface vessels (USVs) – to obtain an Internet of Autonomous Ships (Figure 1). In particular, the system should be able to represent a significant use case in which ICT supports systems and machinery maintenance through sensors and actuators, providing a digital representation of the ship. The ship can be considered a highly complex system composed of heterogeneous and interconnected hardware and software [1]. Managing such a complex system relies on a dense network of thousands of heterogeneous sensors [2] located both onboard and in remote control systems, coupled with a local computing infrastructure.

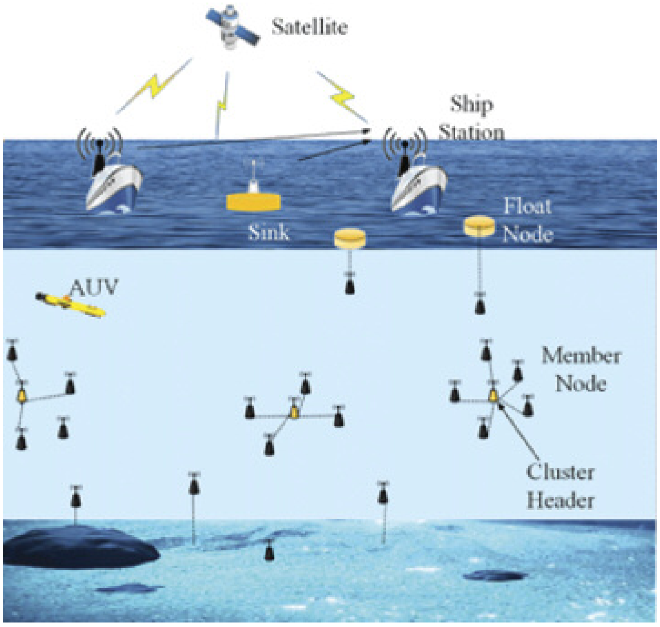


Figure 1: Application scenario.

The local infrastructure can be considered a node of the Internet of Ships (IoS) and its role is to abstract the complexity of the underlying sensor system, resulting in a high-level representation of the management systems, the bridge, and the federated elements (Figure 2). Having a local infrastructure also has indirect benefits, including (i) reduced latency and cost of transmission because the synthesis of useful information, rather than raw data, is transmitted (thus avoiding "data deluge"); and (ii) confidentiality of the data is guaranteed.

Data will be shared within the IoS network to achieve a collaborative aggregation of information and to enable a

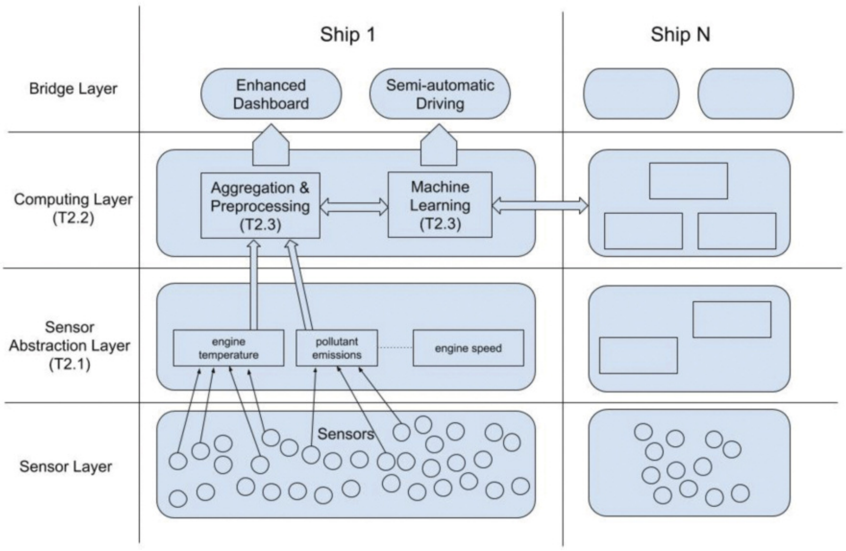


Figure 2: Evaluation process layers.

predictive analysis [3] (such as ship trajectory, drop of the machinery performances, etc.) using artificial intelligence (AI) and machine learning techniques and algorithms under the "federated learning" paradigm. This paradigm relies on a distributed computing system that allows AI mechanisms to be implemented over an extensive network of nodes made up of the autonomous vehicles of a fleet. A single autonomous vehicle (a ship) obtains the current AI model from a remote computing centre (data centre) and subsequently updates its onboard model. The computing centre then summarises the changes made to the AI model and redistributes them to the ships in the network. The data centre's task is, therefore, to merge and standardise all the updates it receives from the ships (network nodes) to improve the global shared AI model, which will thus be updated and made available to all the network nodes. This approach allows machine learning algorithms to be decoupled from data collection and storage, and it requires an appropriate redesign of applications and algorithms. In other words, all training data remains on board the ships, and individual updates are not distributed to other ships or stored in the data centre, ensuring maximum confidentiality of the information (Figure 2). This is crucial when ships from different companies are operating in the same region.

To test the reliability and the feasibility of this complex system, we plan to simulate different operation scenarios for maritime applications. We will test the procedures and algorithms for an unmanned vessel-based smart automation scenario using a self-propelled physical model in scale (which is already in use at the University of Genoa), and many emulated ships through "digital twins" tools (Figure 3). Both real and virtual ships navigate in the same environment and communicate with each other via a data centre, which also manages the AI algorithm. This testbed represents a realistic representation of a crowed navigation area without the needs of multiple vessels, and the system could be tested and debugged with the high safety level.

Our platform offers an innovative way of providing automatic control services and data traffic analysis for the maritime sector, potentially having economic and social benefits for the maritime community. For example, our platform can be useful to analyse operational data to develop a tool capable of guaranteeing a high level of safety also from ashore, moving many services to the ground. In this way, the number o activities performed by the workers in dangerous and stressful conditions can be decreased.

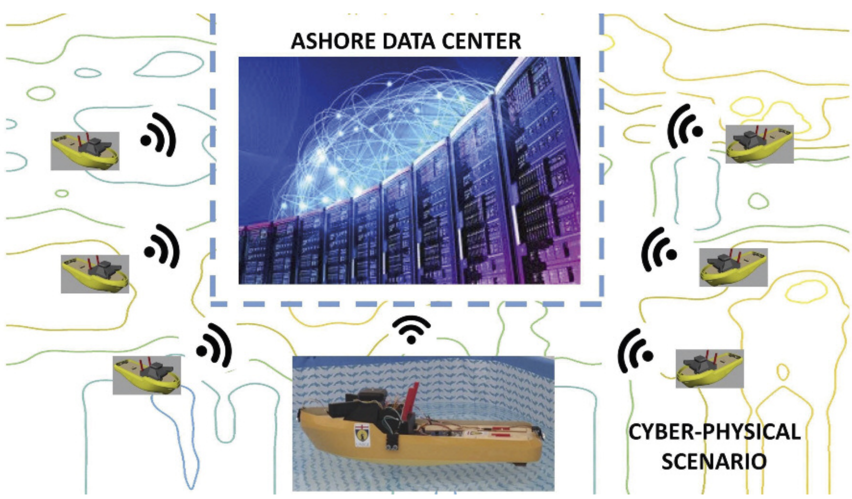


Figure 3: Cyber-physical integration for testing.

References:

[1] M. Martelli, M. Figari: "Real-Time model based design for a CODLAG propulsion control strategies", Journal of Ocean Engineering, vol. 141, pp. 265-276, 2017.

[2] M. Bacco, et al.: "IoT Applications and Services in Space Information Networks", IEEE Wireless Communication Magazine, vol. 26, issue 2, pp. 31-37, 2019.

[3] P. Cassarà et al.: "A Hybrid Cross-Entropy Cognitive-based Algorithm for Resource Allocation in Cloud Environments"; IEEE Int. Conf. SASO, London, pp-1-9, 2014.

Please contact:

Michele Martelli
Università degli Studi di Genova, Italy
michele.martelli@unige.it

Pietro Cassarà
ISTI-CNR, Italy
pietro.cassara@isti.cnr.it

Intelligent Journey Planning with Multi-hop Scheduling Optimisation for the Yachting Sector

by Andreas Komninos (Computer Technology Institute & Press “Diophantos”), Charalampos Kostopoulos (OptionsNet IT Services & Consulting) and John Garofalakis (Computer Technology Institute & Press “Diophantos”)

Yachting tourism has the potential to drive strong blue growth in hosting countries due to its multiplicative effects in other related service and goods economy sectors. The project “Intelligent ICT Applications for the Management of Marinas and Yachts” is using intelligent approaches to address the unique challenge of providing journey-planning tools to yachters.

Yachting is a major contributor to the high-quality, sustainable, blue growth economy in the Mediterranean. In Greece, yachting contributes approximately 4.4 jobs for every 100 docking spaces, and a further 100 jobs in related business sectors [L1], adding tremendous value to the local and national economy. Research and application of digital technology can enable further growth of this sector.

The project “Intelligent ICT Applications for the Management of Marinas and Yachts” (SaMMY+) [L2], is a partnership between industry and academia. The industry partner, OptionsNet, provides digital services to the marina booking and management sector, with its product platform SaMMY [L3]. The academic partner, the Computer Technology Institute and Press “Diophantos”, has expertise in algorithmic solutions for spatio-temporal data, and machine learning. Both partners are based in Patras, Greece. One of aims of the project is to improve the SaMMY platform, with a customer-facing multi-criterion journey planning service.

Route-finding in yachting is harder than for road travel because the connections between docking points (waypoints) in a journey are not fixed in space and time.

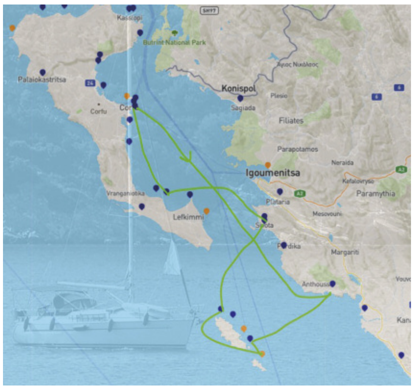


Figure 1: Planning a yachting holiday route is a difficult challenge. A typical 7-day itinerary in the Ionian Sea can be almost any combination between the 35 ports and marinas shown here (Map data © OpenStreetMap, port and marina locations courtesy of http://mykosmos.gr, sample itinerary from http://charter-greece.com)

A vessel can take any course between two waypoints, restricted only by geography and supplies (Figure 1). Variable weather conditions make it difficult to predict how long it might take to travel between two points, or even whether it is possible. Further, yacht travel planning involves queries about amenities at each destination (e.g. docking availability, fuel, nearby facilities), which typically do not pertain to land travel. A yachter aiming to plan a journey must thus: (i) identify candidate waypoints based on availability, on-site and/or nearby amenities, and (ii) calculate an optimal route between waypoints, with optimality depending on several criteria (e.g. maximising stay time at each point, or the number of visited points). This is a difficult cognitive task, requiring significant manual effort. Our aim is to develop algorithms to support the planning process.

To identify candidate waypoints, we employ a semantic approach with docking points and other POIs modelled in an ontology (Figure 2). Compared to a traditional spatial database, an ontology offers much better support for reasoning to handle abstract queries likely to be posed by non-expert users during a search (e.g. “I have five days and want to visit Ithaca, Lefkas and Corfu and like to eat seafood”). It also offers better modelling flexibility (classes and relationships can easily be added or modified) and allows for fuzzy representations of relationships (e.g. “isNearBy” can be a relationship type between two POIs without concern for the actual distance).

To find an optimal route between waypoints, we model travel as a weighted, time-dependent graph, where vertices represent docking points, edge weights represent travel time and stay time can

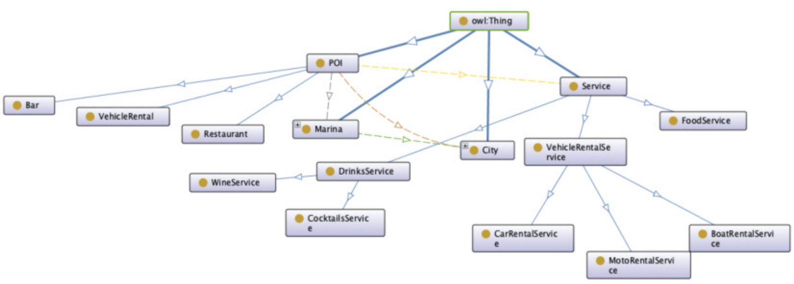


Figure 2: An example ontology for marina candidate determination. Various POIs (including marinas) belong to cities. POIs can be related to various service types, and to marinas through the “isNearMarina” relationship.

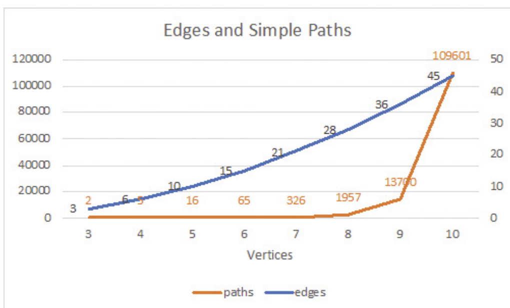


Figure 3: The search space (number of possible simple paths) increases exponentially as more candidate docking points (vertices) are added to the graph.

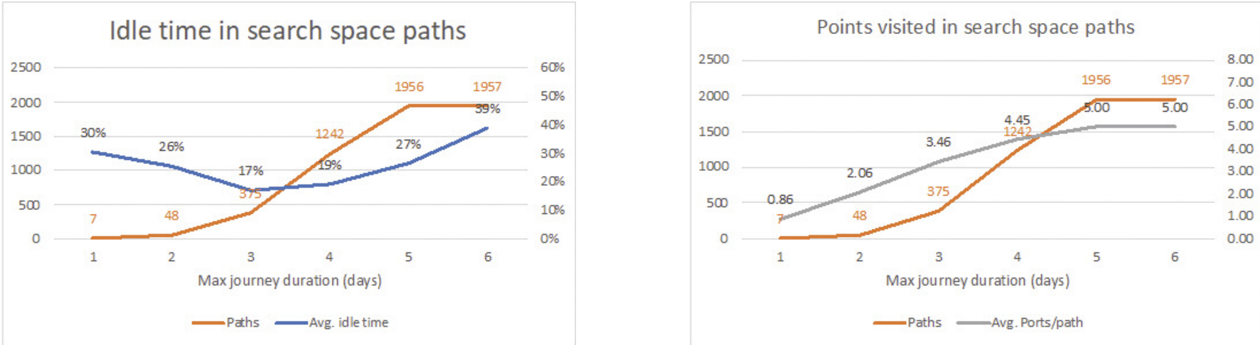


Figure 4: Qualitative aspects of path search space. For our given graph model, it appears that the best compromise is short trips (three days) since on average these trips offer the lowest idle time and the search space is reasonably small. Longer maximum journey durations allow for more ports to be visited (up to five, on average), but most solutions in this search space lead to very large idle times.

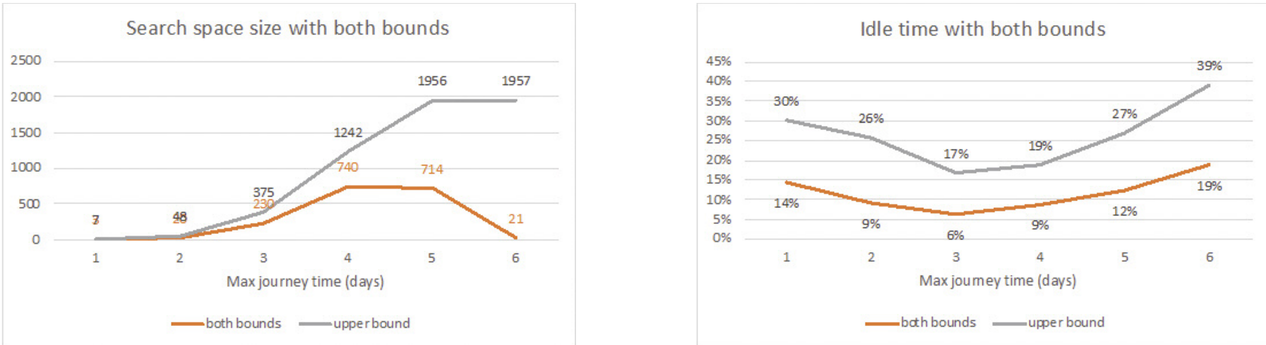


Figure 5: Applying a restriction on the remaining idle time (here, 20%) significantly shortens the path search space, especially for trips with a longer maximum duration in our graph. Though for our given graph the best quality trips (less idle time) are three days, the search process can still return good results even for longer trips.

be included at various vertices [1, 2]. In contrast to land travel, where vertices are typically connected by a single edge, in sea travel every vertex is theoretically connected to all others (within reach of the vessel's supply levels). This can translate to an exponentially large number of possible routes that can be investigated (the "search space"). For example, in a complete non-directional graph (all vertices connected to each other) with 10 vertices, selecting any two vertices as the start and end point, the search space includes 109,601 simple paths (Figure 3). Adding the user's maximum available journey time to the problem can help limit the search space by using a backtracking algorithm, since this approach can quickly eliminate many of the possible paths that lead to a longer journey time [3].

To demonstrate, we use a randomly generated full, non-directional graph with eight vertices and edge weights between 30 and 480 minutes and consider a fixed stay time at each vertex of 720 minutes (12 hours). A shorter maximum journey duration significantly limits the search space in this model (Figure 4, left). However, this creates another problem – we notice that for many of the possible solutions, there remains significant "spare time" (i.e. time after arrival at the end point, and the maximum user-specified journey time, see Figure 4). This means that many of these solutions are low quality, since to reduce this idle time, the yachter must stay longer in each (or one) port, which in turn may affect availability of spaces in subsequent destinations, and so on. Adding a threshold on the maximum allowable idle time (e.g. 20% of the maximum journey time) significantly improves the limitation of the search space (Figure 5). Furthermore, it allows for a better visitation experience, since the possible trips contain, on average, more points than using the maximum journey duration limit alone.

In the near future we aim to assess the performance of these algorithms using a realistic dataset. To populate the POI ontology, we have developed a scraping system to mine data from various sources (e.g. Foursquare, Facebook, Google Places). A problem that remains to be solved is the estimation of realistic travel times between docking points. For this, we aim to mine and analyse historical data from AIS records, according to vessel type and declared origin and destination.

Links:

[L1] https://kwz.me/h1J
[L2] http://hermes.westgate.gr/sammy/
[L3] http://sammyacht.com

References:

[1] H. Wu et al.: "Path problems in temporal graphs", Proc. VLDB Endow 7(9):721–732, 2014.
[2] G. V. Batz, P. Sanders: "Time-dependent route planning with generalized objective functions", in: European symposium on algorithms, Springer, pp 169–180, 2012.
[3] A. Biere, M. Heule, H. van Maaren: "Handbook of Satisfiability", IOS Press, ISBN 978-1-60750-376-7, 2009.

Please contact:

Andreas Komninos, Computer Technology Institute and Press "Diophantos", Greece,
akomninos@ceid.upatras.gr

Charalampos Kostopoulos, OptionsNet IT Services & Consulting, Greece
linos@optionsnet.gr

A Geolocation and Smart Alert System for Nearby First Responders on Roll-on/Roll-off Vessels

by Paschalis Mpeis (University of Cyprus), Jaime Bleye Vicario (Centro Jovellanos, Spain), and Demetrios Zeinalipour-Yazti (University of Cyprus)

A4IoT is an innovative localisation architecture that supports a smart alert system to provide monitoring, navigation and guidance to first responders during fire outbreaks on roll-on/roll-off (Ro/Ro) vessels.

With global trade increasing over the last decade, the maritime industry has grown significantly. Each month, more than 10,000 vessels pass through the Strait of Dover, including cargo ships, tanker ships, roll-on/roll-off ships (Ro/Ro), passenger ships and others. Globalisation and the recent advent of low-cost manufacturing facilities has fuelled an interest in innovative systems to tackle existing issues to lower the costs, improve safety standards and comply with international and regional regulations.

Vessel-owners and passengers have the expectation that, just like other smart spaces (e.g., factories and hospitals), marine vessels will benefit from the latest Internet of Things (IoT) technology. One obvious application of this technology is localisation – obtaining the geographic location of an object by means of digital information processed via a network. Although satellite-based localisation, e.g., GPS, is globally available, it works only in outdoor spaces and is obstructed by the bulky steel structures of vessels. Additionally, vessels lack onboard hardware infrastructure that facilitates indoor localisation, e.g., dense networks of Wi-Fi or UWB, due to the high installation and maintenance costs. On the other hand, there is a growing expectation and a need to localise assets on vessels and to have efficient solutions that work even in the harshest conditions.

The technology that performs accurate on-vessel localisation carries out a variety of important tasks, including: asset tracking, monitoring, analytics, navigation and safety. The localisation literature is very broad and diverse as it exploits several technologies. GPS is ubiquitously available but is energetically expensive and cannot operate in indoor environments. Alternative solutions include: infrared, bluetooth low-energy (BLE), visual or acoustic analysis, RFID, ultra-wide-band (UWB), wireless LANs, or a combination of these into hybrid systems.

In the context of the funded EU Horizon 2020 LASH-FIRE project (nº 814975), which aims to integrate new and

Figure 1: The Fire Resource Management Centre (FRMC) of the Stena Jutlandica Ro/Ro vessel will be augmented with A4IoT to enable real-time tracking of fire hazards and to guide trained personnel in controlling them.

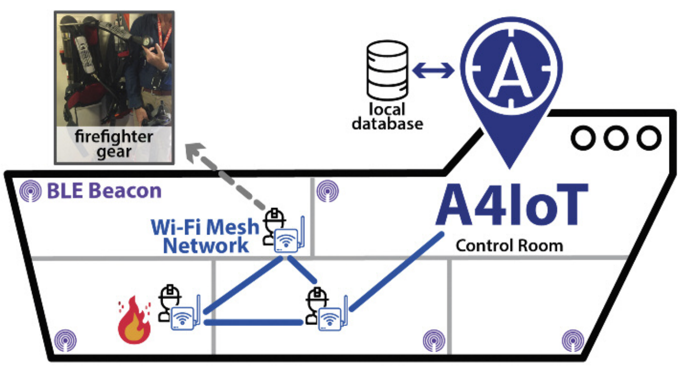


Figure 2: Architectural diagram of A4IoT in a vessel that uses statically attached BLE beacons, and portable IoT devices attached to firefighters' gear, with Wi-Fi mesh-capable routers, BLE receivers, and cameras. Localisation relies on a 2-layer network: a) the Wi-Fi mesh-topology network, that enables connectivity between A4IoT and the firefighters for exchanging text, audio, or video, and b) a BLE beacon sensor broadcast network that will provide accurate localisation within the vessel through RSSI measurements.

advancing technologies to improve safety and security, environmental and personnel impacts and facilitate international trade, we are focusing on effective fire management operations on ro/ro vessels. We are developing a smart alerting system in the form of a vessel indoor information system (see Figure 1) for nearby first responders. The system provides indoor fire intelligence: data collection, data alignment of measurements, activity recognition and orientation, heatmap and crowdsourcing. Our platform will enable messages to be sent (text, audio, video or images) to the crew around the activated fire detector with important safety information about tackling the fire.

Our proposed architecture comprises data, localisation and network layers, as shown in Figure 2. In the control panel of a vessel, an A4IoT backend is automatically deployed using docker [1]. The data layer is kept locally in a document database, the RM data. For the localisation layer we use a long-range beacon-based BLE fingerprint localisation algorithm that provides deck-level accuracy. We are also investigating the use of a computer vision- (CV) based, infrastructure-less localisation algorithm using YOLO [L1] and OpenCV. For the BLE system, we are installing a set of beacons across the vessel's indoor spaces. For the CV technology, we will attach IoT devices with cameras, as well as portable network devices, onto firefighters' gear. This will create a mesh-topology network on-the-fly and give a secure connection between firefighters and the backend, enabling text, audio, video, and localisation data exchange.

To provide localisation, A4IoT [L2] uses a hybrid indoor radio map (RM) that uses the concept of fingerprinting and works seamlessly on the edge/IoT devices. It can operate without connection to the internet with radio signals from different sensors, e.g., Wi-Fi, BLE, or CV. In an offline phase, a logging application records the "fingerprints", which consist of received signal strength indicators (RSSI) of these sensors at certain coordinates (x,y) pinpointed on a vessel's deck map (e.g., every few metres). In the case of CV, it uses extracted textual information from captured images in the place of the RSSI values. Subsequently, in a second offline phase the sensor fingerprints are joined into several N x M matrices, termed the "RM" (i.e., one RM per sensor type), where N is the number of unique (x,y) fingerprints and M the total number of beacons. Finally, an IoT device attached to a firefighter can compare its currently observed fingerprint against the respective sensor-type RM to find the best match, using known algorithms such as KNN or WKNN [2]. Our solution can be infrastructure-free, as it performs best-effort operation with whatever sensor type is available. This information can then be used to guide the firefighters.

Links:
[L1] http://pjreddie.com/darknet/yolo/
[L2] https://anyplace.cs.ucy.ac.cy

References:

[1] P. Mpeis et al.: "The Anyplace 4.0 IoT Localization Architecture", in 2020 21st IEEE International Conference on Mobile Data Management (MDM), pp. 218-225, IEEE, 2020.
[2] C. Rizos et al.: "Indoor positioning techniques based on wireless LAN," in 1st IEEE Intl. Conf. on Wireless Broadband and Ultra Wideband Communications. IEEE, 2007, pp. 13–16.

Please contact:

Demetrios Zeinalipour-Yazti
Data Management Systems Laboratory (DMSL), University of Cyprus
+357 22 892755, dzeina@cs.ucy.ac.cy

Understanding Complex Attacks on a Maritime Port

by Sandra König (AIT Austrian Institute of Technology)

Attacks on maritime ports have become more sophisticated since modern ports turned into cyber-physical systems. Simulation models can help with the vital task of detecting such attacks and understanding their impacts.

Digitalisation introduces new challenges to the protection of modern critical infrastructures, such as maritime ports. While control systems ensure smooth physical operations, they are also accompanied by new threats. Complex attacks such as advanced persistent threats (APTs) or drug smuggling [L1] make explicit use of the interconnection between cyber and physical systems of a port. Their stealth makes them difficult to detect and their potential impacts can only be estimated. Impact estimations in this context should be based on a formal analysis of the system. During the course of the European Commissions project SAURON [L2] a model has been developed that simulates the aftermath of a security incident in a port in order to understand the impact on the port as well as the local population.

The simulation model represents the maritime port as a graph where nodes describe relevant assets, and edges describe a dependency between two assets. Assets can be physical (crane, gate, truck or camera), cyber (server, working laptop or a database), but also represent processes (identification of employees, registration of a container). Once the dependency graph is known, the internal dynamics of each asset are modelled. Its functionality is described on a three-tier scale, where states can be interpreted as "working properly" (state 1), "partially affected" (state 2) or "not working" (state 3). The state of the asset changes depending on notifications about events that have happened. Due to the complexity of the considered attacks, the state changes are assumed to happen with a certain likelihood. Once these likelihoods are determined, it is possible to mimic how an attack spreads through the entire system. A formal description of the model is given in [1] and the idea is illustrated in Figure 1.

A practical application of the formal model to a concrete problem follows these steps [2]:

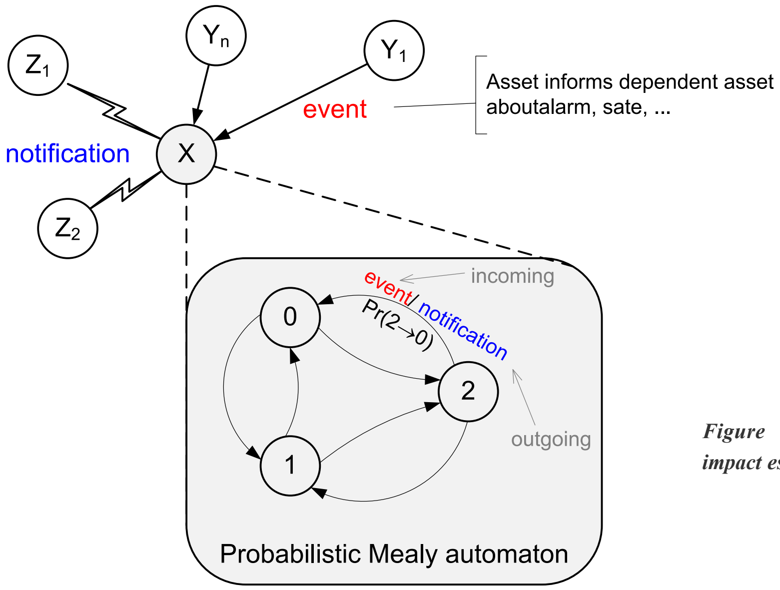


Figure 1: Simulation model for impact estimation.

- Identification of relevant assets. This step answers the question “Which components are important for smooth operation?” and the answer is a list of assets containing, for example, server, computer, camera, truck etc. If the number of assets is high, it is beneficial to classify them to reduce the modelling effort later by labelling classes instead of individual assets.
- Identification of dependencies. This step answers the question “How do the identified assets depend on one another?” and the answer is a list of pairs (u,v) where v depends on u. There are many ways that dependencies may occur, ranging from physical proximity to required services (e.g. identification of employees requires a properly working database). A dependency graph illustrates the identified relationships nicely.
- Identification of relevant alarms. This step answers the question “What threatens the system?” and the answer is a list of alarms. Not every alarm is dangerous for each asset (e.g. a malware attack only affects cyber assets) but the list is a collection of all incidents that reduce functionality of even a part of the port. Each alarm carries information on its kind (e.g., fire, malware, …) and criticality (measured on a fix scale). During simulation, it is also important to attach a timestamp indicating when the incident happened.
- Description of internal dynamics. This step answers the question “How do we describe each asset’s behaviour?” and the answer is a set of states (e.g., ranging from 1 to 3) and for each asset a matrix of transition likelihoods that describe the change between the states. If enough data is available, machine learning techniques such as logistic regression can be applied to learn thse probabilities from expert opinions.

After performing all these steps, the consequences of a concrete incident can be simulated. The considered incident may affect a single asset (e.g., in case of a malware that is activated on a single laptop) or several assets simultaneously (e.g., hackers may switch off cameras). An online tool [L3] is available to perform the significant number of simulation rungs required for statistical analysis.

The simulation allows the impact on assets to be statistically analysed. It is possible, for instance, to estimate how often a specific asset is not working properly (state 2 or 3) or not working (state 3). Further, the average state of each asset can be computed and the corresponding nodes in the dependency graphs can be coloured accordingly (using green, orange and red to indicate the average level of functionality). The resulting graph gives a quick overview of which assets are threatened the most. Another important result of the analysis is that it allows the cause of failure of a specific asset to be identified by following back the chain of infection that led to the failure. Such information is particularly useful when investigating ways to protect the system – an asset that causes problems should be replaced or protected to increase the security of the entire system.

This work was supported by the European Commission’s Project SAURON (Scalable multidimensional situation awareness solution for protecting European ports) under the HORIZON 2020 Framework (Grant No. 740477).

Links:
[L1] https://kwz.me/h2m
[L2] https://www.sauronproject.eu/
[L3] https://atlas.ait.ac.at/sauron/

References:
[1] S. König, S. Rass, B. Rainer and S. Schauer, “Hybrid Dependencies Between Cyber and Physical Systems”, in Advances in Intelligent Systems and Computing, Proceedings of the 2019 Computing Conference, Volume 2, London, UK, July 16-17, 2019, K. Arai Ed. Cham: Springer, 2019. pp 550-565.

[2] S. König, A. Gouglidis, S. Rass, N. Adams, P. Smith and D. Hutchison, “Analysing Disaster-Induced Cascading Effects in Hybrid Critical Infrastructures: A Practical Approach”, in Guide to Disaster-Resilient Communication Networks, J. Rak, D. Hutchison, Eds. Cham: Springer, 2020. pp. 769-789.

Please contact:
Sandra König
AIT Austrian Institute of Technology, Austria
sandra.koenig@ait.ac.at

Computational Approaches for Uncertainty Quantification of Naval Engineering Problems

by Riccardo Broglia, Matteo Diez (CNR-INM) and Lorenzo Tamellini (CNR-IMATI)

The design of efficient seagoing vessels is key to a sustainable blue growth. Computer simulations are routinely used to explore different designs, but a reliable analysis must take into account the unavoidable uncertainties that are intrinsic to the maritime environment. We investigated two ways of performing this analysis in an effective, CPU-time parsimonious way.

The sustainable growth of the marine and maritime sectors involves ship designers, shipyards and operators in the quest for energy-efficient ships and safe operations at sea. Computational models, such as computational fluid and structural dynamics, can provide accurate predictions of complex physical phenomena and, combined with design/operational space exploration methods, drive the decision-making process.

However, even assuming perfect models and error-free solvers, multiple sources of uncertainty (such as variability of sea and operational conditions) impact the performance analyses. Ship design and performance prediction must take into account this uncertainty to provide meaningful information to decision-makers. For instance, the goal might be to design a ship that displays optimal resistance and seakeeping on average for a wide range of sea conditions, or alternatively to optimise worst-case scenario performance.

Such results can only be obtained by first testing several scenarios, and then computing statistical indicators such as expected value, variance, quantiles, that summarise the variability of the quantities of interest (i.e. the outputs of the analysis, such as ship resistance, motion, safety-related quantities). In other words, an uncertainty quantification analysis must be performed and possibly repeated for multiple ship designs until a satisfactory design is found.

Even with today's computers, uncertainty quantification analyses of complex problems are extremely computationally intensive. Performing them in an effective, CPU-time parsimonious way is key to a successful design process – a challenge that we are addressing in a collaboration between CNR institutes: CNR-INM [L1] and CNR-IMATI [L2]. CNR-INM has a long tradition in computational methods for ship performance assessment (as well as in experimental validation, thanks to large experimental facilities), while CNR-IMATI is an applied mathematics institute. A first set of results is reported in [1], where we consider the problem of computing expected value, variance and probability density function of the ship resistance of a roll-on/roll-off passenger ferry, subject to two uncertain operational parameters: ship speed and draught, which are representative of operating conditions including payload. The design of the ferry, shown in panel (a) of Figure 1, is being made available in the framework of the H2020 EU Project Holiship [L3]. More complex problems with a larger number of uncertainties are subject to ongoing investigation.

Two ingredients required for a successful uncertainty quantification analysis are an effective strategy to decide which scenarios should be tested (i.e. which values of ship speed and draught), and an efficient solver to evaluate each scenario. For the latter, we employ X-navis, a multi-grid RANS code developed at CNR-INM (panel (a) of Figure 1 shows the free surface elevation computed by X-navis). The multi-grid nature of the solver is also crucial for the former point, i.e., for the strategy that decides the scenarios to be tested. Indeed, we employ a multi-fidelity strategy: we first solve a substantial number of scenarios on a coarse RANS grid (hence with a limited computational cost) to get a rough estimate of the variability of the ship's resistance over the range of considered values of ship speed and draught, and then iteratively refine the RANS grid and further solve a few additional scenarios to get a more precise estimate for some critical combinations of ship speed and draught. In [1] we employed four grids, ranging from 5.5 million grid cells for the finest mesh down to 11,000 for the coarsest. In addition to the multi-fidelity strategy, we also employed an adaptive paradigm, where the scenarios to be solved for

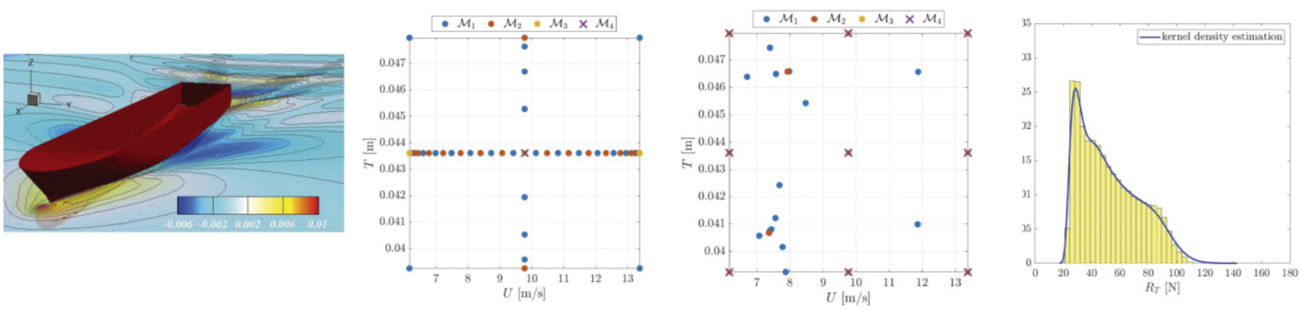


Figure 1: From left to right: (a) the ferry used for the numerical tests, and free surface waves computed by X-navis; (b) and (c) the sampling of the ship speed/draught space provided by MISC and SRBF strategies respectively (i.e. the scenarios solved by the methods); (d) the resulting probability density function of the ship-resistance (we report the results obtained by SRBF only, as those obtained by MISC are comparable). Source: [1]

each RANS grid are not determined a-priori, but decided on-the-go as the computation proceeds, based on suitable criteria.

Two different non-intrusive approaches to uncertainty quantification, resulting in two batches of simulations, have been considered in [1]: the Multi-Index Stochastic Collocation (MISC) method [2] and an adaptive Stochastic Radial Basis Function (SRBF) method [3]. The former samples the ship speed/draught space in a rather structured way (panel (b) of Figure 1), while the latter results in a more scattered sampling (panel (c) of Figure 1). Different colours denote the different RANS grids used at each ship speed/draught combination. Despite the differences in the sampling strategies, both methods give similar results: the average (expected value) ship resistance (at model scale) is around 52 Newton, while the standard deviation is around 22 Newton. The full probability density function, shown in panel (d) of Figure 1, features a heavy tail towards large resistances. Our experience suggests that MISC seems to reach a good accuracy for slightly smaller computational costs than SRBF, while SRBF seems to be more robust to the numerical noise that affects the RANS simulations, especially on the coarser grids. All in all, however, both methods seem to be viable candidates to tackle more complex problems with a larger number of uncertain parameters.

In summary, the design of efficient seagoing vessels is vital for a sustainable blue growth. Ship performance analyses become more meaningful if they report the uncertainty quantification analysis, i.e. the variability of the performance due to the unavoidable uncertainties in the operating regime, which means that several scenarios should be tested, and results condensed in statistical terms (average, standard deviation, etc). Recently developed multi-fidelity, adaptive algorithms can provide these figures in a time-effective way, and further research in this area is mandatory to make these methods even more reliable, effective and ready for the general public.

Links:
[L1] http://www.inm.cnr.it/
[L2] http://www.imati.cnr.it/
[L3] http://www.holiship.eu/

References:

[1] C. Piazzola et al.: "Uncertainty Quantification of Ship Resistance via Multi-Index Stochastic Collocation and Radial Basis Function Surrogates: A Comparison", Proc. of the AIAA Aviation Forum 2020. Also available as arXiv e-prints, 2005.07405.
[2] J. Beck et al.: "IGA-based Multi-Index Stochastic Collocation for random PDEs on arbitrary domains", Computer Methods in Applied Mechanics and Engineering, 2019.
[3] J. Wackers et al.: "Adaptive N-Fidelity Metamodels for Noisy CFD Data", Proc. of the AIAA Aviation Forum 2020.

Please contact:

Lorenzo Tamellini, CNR-IMATI, Italy, tamellini@imati.cnr.it

Matteo Diez, CNR-INM, Italy, matteo.diez@cnr.it

A New Approach to Underwater Technologies for Innovative, Low-cost Ocean Observation: NAUTILOS

by Gabriele Pieri (ISTI-CNR)

The NAUTILOS project aims to fill some of the existing gaps in marine observation and modelling by improving the measurement of chemical, biological and deep ocean physics variables. A new generation of cost-effective sensors and samplers is being developed, integrated into observation platforms and deployed in large-scale demonstrations off the coastline of Europe. These will complement and expand existing observation tools and services, allowing researchers to obtain data at a much higher spatial resolution, temporal regularity and length than is currently available at a European scale. It will also facilitate and democratise the monitoring of the marine environment for both traditional and non-traditional data users.

Environmental observational research in coastal and shelf sea environments, which are characterised by substantial dynamic variability [1], has its own specific challenges. Dense observation efforts in time and space are required to assess biological, chemical and physical processes that are very patchy in nature, such as mesoscale features. Typically, the processes are also highly variable in time due to short-term environmental forcings like upwelling and mixing events that may result in potentially harmful algal blooms. We need a high level of spatial resolution and both high temporal resolution and length (i.e. time-series) to accurately quantify variables such as temperature, momentum, biological and biogeochemical fluxes, how they are changing, and what processes are forcing these changes. High quality data can help us manage and mitigate events that might have adverse climatic, environmental and economic outcomes.

The NAUTILOS project focuses on both the deep ocean at depths of greater than 200 m, and the open ocean environment above it. These are the least observed but largest habitats on our planet (see Figure 1). Future deep ocean observing needs greater integration across disciplines and sectors, best achieved through demonstration projects and facilitated reuse and repurposing of existing deep-sea data efforts [2].

The next step in the evolution of the marine monitoring system will be the widespread adoption of autonomous in

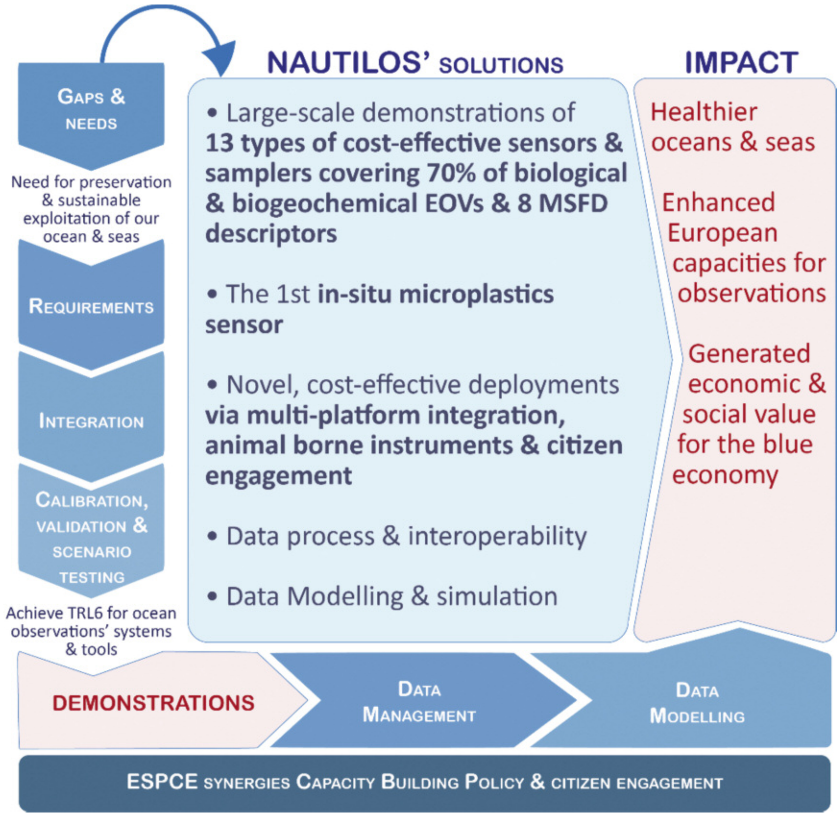


Figure 1: The NAUTILOS Project concept.

situ sensing. There is a need to explore and test new technology that will make it more affordable to acquire, deploy and maintain monitoring and observing stations to fill the in situ observational gaps of ocean observation systems. With increased spatial resolution and frequency of in situ collection and a greater volume of information coming from cost-effective and more widely disseminated sensors, many new operational products will become available. With these will come new forecasting capabilities and processes for understanding the ocean. The improvement and effectiveness provided by this new "digital ocean" observation paradigm based on big data can be statistically evaluated using oceanographic models of different data availability scenarios and evaluating the improvement of forecasting capabilities. These scenarios would reflect the potential increase in spatial resolution, time frame and gap closure of the acquired data anticipated by this project. Technology maturation will promote the “democratisation” of marine environmental monitoring through the production of small, cost-effective sensors.

Such a realisation will broaden the range of users who will benefit from the increase in data and forecasting models. It will enable both traditional users, such as oceanographers, researchers and technology providers, and others, such as citizen scientists, resource managers, policymakers and environmental monitoring groups, to access new and improved resources. The full adoption of the technologies developed and demonstrated in project NAUTILOS can only increase participation in environmental observation, help reduce the costs of the technology, and thus multiply the social, economic and environmental benefits and positive impact of the project.

The NAUTILOS project, funded by the EU under the call H2020-BG-2020-1, is an Innovation Action Pilot that is just starting its activities, which are foreseen to last for four years. The consortium, composed of 21 partners from 11 countries from across Europe, is coordinated by the National Research Council of Italy (participating with three different institutes), has a wide, complementary and multidisciplinary expertise in various fields ranging from oceanography to biology and biogeochemistry, as well as from data processing and modelling to marine technological applications and research infrastructures. Apart from the coordinator , the involved partners are: the “Hellenic Centre for Marine Research”, the “Norsk Institutt for Vannforskning”, the “Suomen Ymparistokeskus”, the “Institut Francais de Recherche pour l'Exploitation de la Mer”, the “Centre National de la Recherche Scientifique”, “ETT S.p.A.”, “EdgeLab S.r.l.”, the “Universidade do Algarve”, “NKE Instrumentation S.a.r.l.”, “Aquatec Group limited”, “SubCTech GMBH”, the “Centro de Engenharia e Desenvolvimento”, the “Haute Ecole Specialisee de Suisse Occidentale”, the “Centre Suisse d'Electronique et de Microtechnique SA”, the “University of Lubljana”, “Fudaçao EurOcean”, the “Deutsches Forschungszentrum fur Kunstliche Intelligenz GMBH”, the “University of Calabria”, the “Instituto do Mar”, and “EvroProject OOD”.

References:

[1] K. H. Wiltshire et al.: “Resilience of North Sea phytoplankton spring bloom dynamics: An analysis of long-term data at Helgoland Roads”, Limnology and Oceanography, 53(4), 1294-1302, 2008.
[2] L. A. Levin et al.: “Global observing needs in the deep ocean” Frontiers in Marine Science, 6, 241, 2019.

Please contact:

Gabriele Pieri, ISTI-CNR, Italy
Gabriele.Pieri@isti.cnr.it

Sea Monitoring Made Simple and Efficient

by Serena Berretta (University of Genova), Daniela Cabiddu (CNR-Imati), Michela Mortara (CNR-Imati) and Michela Spagnuolo (CNR-Imati)

Remote sensing provides almost global spatial coverage, but with limits in resolution and accuracy. Ground stations, conversely, provide very accurate coverage with high temporal resolution, but sparse and pointwise. What's missing from this picture is accurate local knowledge with a high spatial resolution, making it possible to understand and analyse local phenomena to their full spatial extent. To fill the gap, we propose a paradigm shift in field sampling; environmental surveys that are dramatically cheaper, quicker and easier to perform, and the ability to perform visualisation and analysis tasks as the survey progresses. This is made possible by a real-time adaptive sampling method, embedded within a mobile observation platform. The method continuously guides and refines samples acquisition while in the field, elaborating the variable distribution and its related uncertainty along with the sampling process.

Environmental monitoring is crucial for investigating the condition of soil, air and water, describing natural phenomena, and promptly reacting to, and even preventing, accidents. In standard environmental surveys, the sampling locations are predetermined either on a regular grid or are defined by a-priori knowledge. Once collected, samples are subject to laboratory analysis to provide accurate measurements at sample locations and generate geo-statistical maps to represent an estimation of the continuous distribution of the environmental variables over the domain. Further surveys might then be required on subareas of the global domain to reach the desired reliability. Apart from the cost of sampling, the time required for laboratory analysis might preclude a prompt reaction during a critical event.

New technology enables more efficient approaches to sampling: new, accurate dynamic positioning systems coupled with lighter and cheaper sensors can be used as mobile laboratories, which in the near future will yield an explosion of georeferenced, highly accurate in-situ data. This opens the door to on-the-fly sampling decisions – quite the opposite to traditional approaches based on predefined locations.

These new methods require innovative computational solutions to make data analysis precise and fast. To this end, we propose a new sampling methodology, adaptive sampling, based on a fast and effective iterative refinement of geo-statistical maps estimating the distribution of acquired environmental variables. In our setting, environmental variables are measured by sensors mounted on mobile platforms, and the data acquisition step is interleaved with the computation of the distribution map, which is iteratively refined. The concept of uncertainty of the estimation plays a key role in the refinement process and switches from a "passive" role in traditional sampling, where it was used retrospectively to measure the accuracy of the final model, to an "active" role, by intervening in the definition of "smart" sampling locations. The new sampling site is located where we are less confident in our prediction, i.e., in locations of maximum uncertainty.

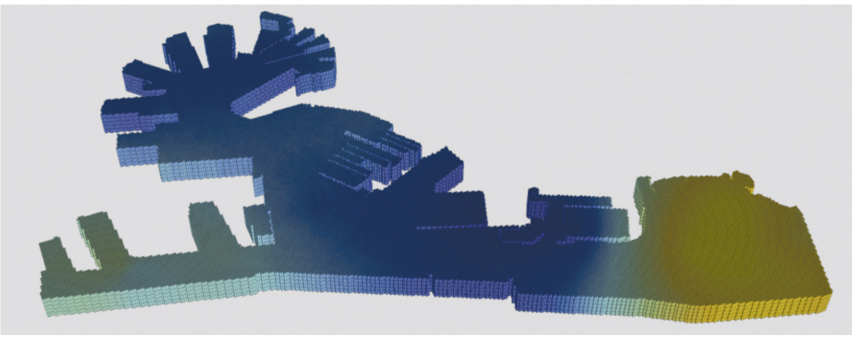


Figure 1: A geometric representation of the harbour in Genoa (Italy). It represents the water volume within the pier structures and the seabed. The 3D model is a structured grid of voxels of equal size and it is generated starting from bathymetric data and known boundaries of the piers. The different colours show an example distribution of an environmental variable associated to grid cells.

To make this happen we need three main ingredients: (i) the geometric representation of the spatial domain of the environmental variables; (ii) physical samples acquired by modern accurate sensors; (iii) geo-statistical and computational tools to analyse and correlate data, and to estimate values in unknown locations and their relative uncertainty.

To make our adaptive sampling work, the digital geometric representation of the survey area is built as a volumetric model, a structured grid of voxel/cells of equal size, while the acquired raw data undergo filtering, synchronisation and interpolation as pre-processing before entering the reconstruction cycle. Then, the prediction and the uncertainty maps are computed on the volumetric grid, which results in the predicted variable distribution map and a new suggested sampling position. The computation of the prediction and uncertainty maps is performed at each iteration, that is, every time a waypoint is reached. All acquired samples contribute to identify the covariance law among data, that is, intuitively, how similar we expect two measures to be based on their spatial distance. This covariance law, expressed by the function variogram [1], allows us to predict values in unsampled locations by running several Gaussian Simulations [1] and by assigning both an estimated value (the mean of the simulation outputs) and an uncertainty value (the variance of the simulation outputs) to each cell.

As a first case study, we are focussing on water monitoring in the harbours of Genova (Italy) (Figure 1) and Toulon

(France). Anyway, the system has the potential to be included in the monitoring practices of other harbours and additional case studies are currently under definition. Preliminary results [2] show that our adaptive sampling can reach the same estimated accuracy as traditional approaches do, but with fewer samples. We believe that this approach can therefore facilitate faster, cheaper, highly efficient monitoring systems, ideal for providing real-time environmental variable estimates in emergency scenarios. Mobile sensor platforms allow users to bypass laboratory analysis, making the survey more affordable, thus enabling a higher sampling frequency.

Future work will focus on unstructured grid representation of the water volume to better align with the complex boundaries of the piers (e.g. using tetrahedral cells). Since cells will have uneven volumes, a proper change of support strategy must be planned first; this is an open problem in mathematical research that we are currently studying. Finally, by changing the robotic platform (e.g. to drones), the approach is easily transferable to the environmental monitoring of soil and air.

This work was conducted within the framework of the European project “Real-time Adaptive Monitoring with Automation of Sampling - Port Coastal Areas - MATRAC-ACP”, funded by the Interreg Italy-France Maritime 2014-2020 Program - Axis priority 2, Specific objective 6C2 “To increase protection of marine waters in ports”.

Link:
[L1] http://interreg-maritime.eu/web/matracacp

References:
[1] Chiles, J. P., & Delfiner, P. (2012). Geostatistics: Modeling Spatial Uncertainty. 2-nd ed.-NewYork, Toronto, JohnWiley'Sons.
[2] Berretta, S., Cabiddu, D., Pittaluga, S., Mortara, M., Spagnuolo, M., & Zuccolini, M. V. (2018). Adaptive Environmental Sampling: The Interplay Between Geostatistics and Geometry. In STAG (pp. 133-140).

Please contact:
Serena Berretta
University of Genova, Italy
serena.berretta@ge.imati.cnr.it

Offshore Oil & Gas Blow-out Containment

by Fivos Andritsos

The demand for hydrocarbons and other mineral resources worldwide is increasingly met by tapping the vast seabed resources in ever more difficult and risky environments, like the abyssal oceanic depths or the arctic regions. The development of suitable means for safe underwater operations is a fundamental requisite for their sustainable exploitation, in particular on the delicate polar environments or the environmentally and socially stressed east Mediterranean region.

Maritime disasters leading to major pollution happen almost regularly every 2-3 years. However, it is the offshore drilling well blow-outs that have caused the major environmental disasters, like the IXTOC-1 in 1979 or the DEEPWATER HORIZON in 2010, in the Gulf of Mexico. Given that most of the new offshore hydrocarbon deposits are found under thick mineral salt layers, often several km below the sea-floor, the associated risks are very high. Hence, the community has to ensure that all necessary prevention and emergency intervention means are in place, especially in the sensitive arctic environment or in seas with densely populated, touristic coastlines like the eastern Mediterranean.

In the aftermath of the “PRESTIGE” oil-tanker disaster, a novel method for the containment of the pollutants directly on the shipwrecks, even at abyssal depths, was conceived by the author and patented by the European Commission. It consists of a large fabric dome (collector), solidly anchored at the sea bed, so as to cover entirely the pollution source and channel all the leaking oil, through a wide vertical tube (riser) to a large open bell-shaped reservoir (buffer bell) positioned 10-20 m under the sea surface so as not to be affected by the waves. Oil occupies the upper part of the bell and is periodically removed by a shuttle tanker while water escapes from the open bottom. The buoyancy of the reservoir keeps the whole system in tension.

The method, schematically depicted in Figure 1, is simple, entirely passive and weather independent. Extensive numerical simulations and scale experiments, in the frame the DIFIS (Double Inverted Funnel for Interventions on Shipwrecks) EC-financed collaborative project [L1], provided the proof of concept, the proper engineering solutions (Figure 2) as well as the technical and economic feasibility for interventions on ship wrecks lying on the seabed from few hundred meters till several km deep [1-3].

The “DEEPWATER HORIZON” catastrophe triggered additional investigations on the applicability of the DIFIS method for containing deep offshore well blowouts. The main concern was the presence of gas that, under certain conditions, could compromise the stability of the system. Extensive 3-phase flow numerical simulations, using a variety of commercial and home-made codes, indicated that the presence of gas in the upward flow, even in significant quantities, would not be a matter of concern as long as the gas accumulated in the “buffer bell” is continuously evacuated towards the surface.

Concluding, the DIFIS system is by far the best way to intervene also on blown-out offshore wells in order to contain the pollutants right on the seafloor before they disperse on the seawater volume and reach the sea surface, where it is extremely difficult to deal with. It offers a range of advantages over the current state-of-art:

- Flexibility: it can be applied on a variety of underwater oil pollution sources like ship-wrecks or blown-

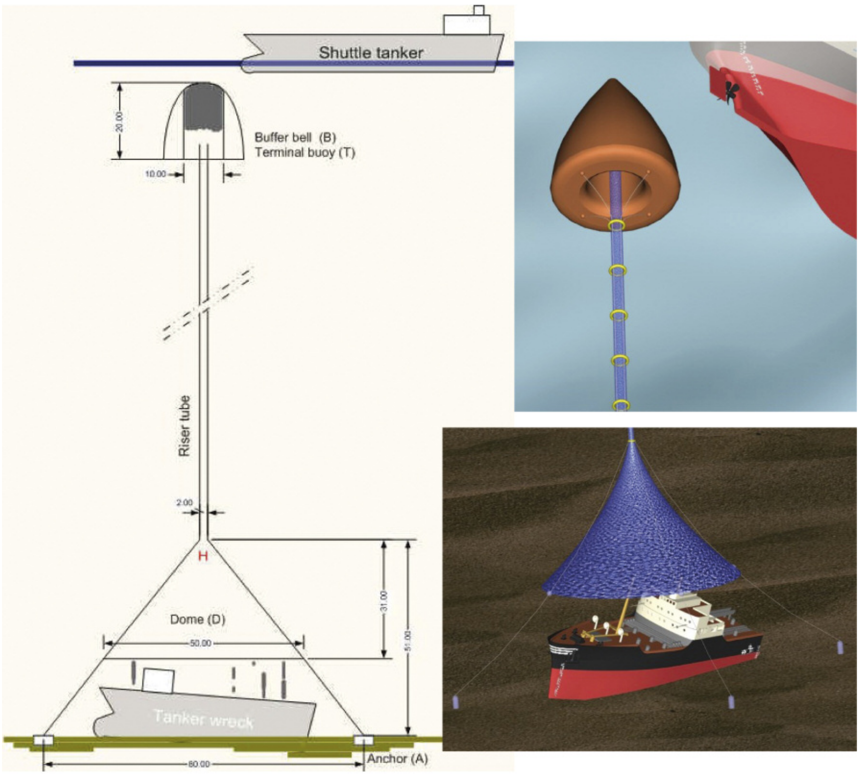


Figure 1: Schematic layout (left) and initial 3-D models of DIFIS system.

out offshore wells, from 400 m to more than 4,000 m deep.
- Does not interfere with the blow out preventer (BOP) or wellhead structure.
- Simplicity: entirely passive, "place & forget" design, once in place, no remotely operated vehicles (ROV) or mother ship required (excl. offloading).
- Rapid deployment: components and deployment tool modules can be air-lifted and assembled in place. For depths less than 2,000 m, underwater deployment can be done through an inexpensive, purpose-built side-derrick attached to any vessel with dynamic positioning capacity.
- Rough weather tolerant: once installed, the only surface operations concern the periodic offloading of the buffer bell reservoir.
- The large riser tube, coupled with the one-step deployment procedure ensure the system against any blockages due to hydrate formation.
- Blown-out gas either forms solid hydrates, dissolves in the water or expands upwards breaking down in small bubbles. Resulting dynamic forces are orders of magnitude less than the tensile buoyancy forces so that in no way can they compromise the structural stability of the system.

In the case of offshore installations, the speedy and cost-effective deployment of the DIFIS system can be ensured by prefabricating some standard components and by building some limited underwater infrastructure, like the anchoring points, together with the other seabed production installations.

Link:
[L1] https://kwz.me/h2C

References:

[1] F. Andritsos, H. Cozijn: "An Innovative Oil Pollution Containment Method for Shipwrecks proposed for Offshore Well Blowouts", Proceedings of the ASME 2011 30th International Conference on Ocean, Offshore and Arctic Engineering (OMAE 2011), June 19-24, 2011, Rotterdam, The Netherlands.

[2] P. Sotiropoulos, N. Tosi, F. Andritsos, F. Geffard: "Optimal Docking Pose and Tactile Hook-Localisation Strategy for AUV Intervention: The DIFIS Deployment Case", Ocean Engineering 46 (2012) pp 33–45, Elsevier.

[3] D.E. Mazarakos, F. Andritsos, V. Kostopoulos: "Recovery of Oil-Pollutant from Shipwrecks: DIFIS Project", International Journal of Structural Integrity (ISSN 1757-9864), Vol. 3 Issue: 3, pp.285-319.

Please contact:

Fivos Andritsos
Mechanical Engineer (PhD)
fivos.andritsos@gmail.com

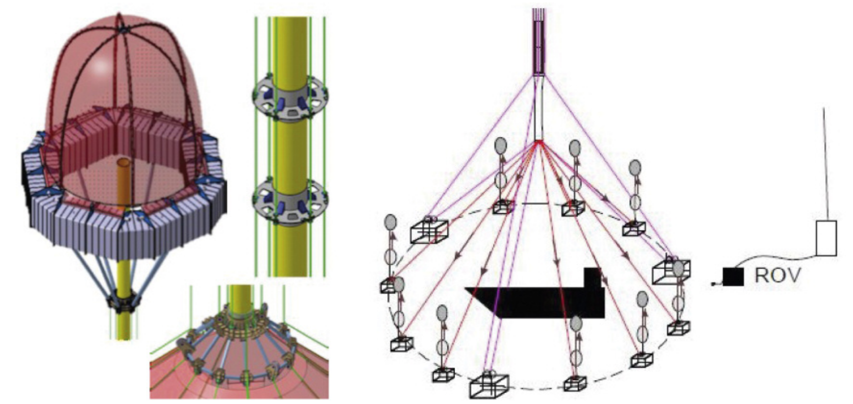


Figure 2: Design details of the buffer bell, the riser tube and its interface to the collecting dome (left) and the underwater deployment procedure (right).

Figure 3: DIFIS project partners.

Evaluation of Synthetic Data for Privacy-Preserving Machine Learning

by Markus Hittmeir, Andreas Ekelhart and Rudolf Mayer (SBA Research)

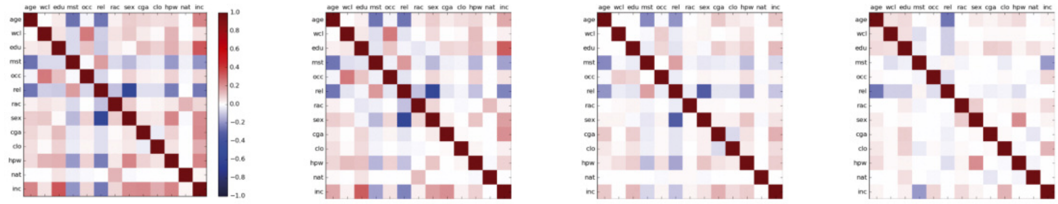
The generation of synthetic data is widely considered to be an effective way of ensuring privacy and reducing the risk of disclosing sensitive information in micro-data. We analysed these risks and the utility of synthetic data for machine learning tasks. Our results demonstrate the suitability of this approach for privacy-preserving data publishing.

Recent technological advances have led to an increase in the collection and storage of large amounts of data. Micro-data, i.e. data that contains information at the level of individual respondents, is collected in domains such as healthcare, employment and social media. Its release and distribution, however, bears the risk of compromising the confidentiality of sensitive information and the privacy of affected individuals. To comply with ethical and legal standards, such as the EU's General Directive on Data Protection (GDPR), data holders and data providers have to take measures to prevent attackers from acquiring sensitive information from the released data.

Traditional approaches to compliance often include anonymisation of data before publishing or processing, such as using k-anonymity or differential privacy. Synthetic data offers an alternative solution. The process of generating synthetic data, i.e. data synthetisation, generally comprises the following steps:

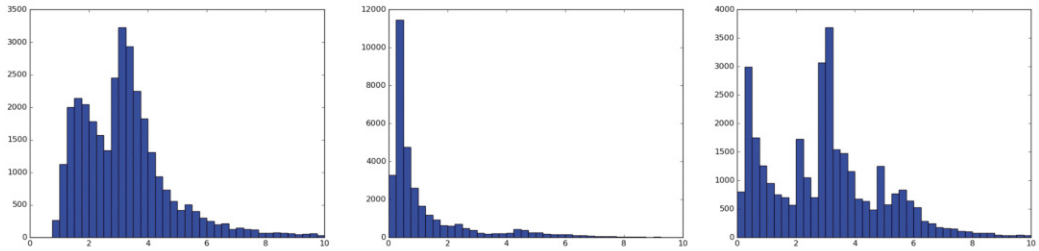
- Data description: The original data is used to build a model comprising information about the distribution of attributes and correlations between them.
- Data generation: This model is then used to generate data samples. The global properties of the resulting synthetic dataset are similar to the original, but the samples do not represent real individuals.

The goal of this technique is that analysis methods trained on the synthetic instead of the real data do not perform (notably) worse. The use of synthetic data should also reduce the risk of disclosure of sensitive information, as the artificially generated records do not relate to individuals in the original data in a one-to-one correspondence. Consequently, validating the utility and privacy aspects is crucial for trust in this method. We conducted an empirical evaluation, including three open-source solutions: the SyntheticDataVault (SDV) [L1], DataSynthesizer (DS) [L2] and synthpop (SP) [L3]. The SyntheticDataVault builds a model based on estimates of the distributions of each column. Correlations between attributes are learned from the covariance matrix of the original data. The model of the DataSynthesizer is based on a Bayesian network and uses the framework of differential pri-



(a) Original (b) DS (c) SDV (d) DS with DP

Figure 1: Heatmaps for SyntheticDataVault and DataSynthesizer on the Adult Census Income dataset [L4].



(a) SDV (b) DS (c) DS with DP

Figure 2: Distance Plots for SyntheticDataVault and DataSynthesizer on the Adult Census Income dataset.

vacy. Finally, synthpop uses a classification and regression tree (CART) in its standard settings.

The utility of the generated synthetic data can be assessed by evaluating the effectiveness of machine learning tasks. Models that are trained on the synthetic data can be compared with models trained on the original data, and scored on criteria such as accuracy and F-score for classification problems. We studied classification [1] and regression [2] tasks on publicly available benchmark datasets. While the results vary depending on the number of attributes, the size of the dataset and the task itself, we can identify several trends. In general, models based on synthetic data can reach utility up to or very close to the original data. Models trained on data from the DataSynthesizer without Differential Privacy or on data from synthpop with standard settings tend to achieve utility scores that are close to those of the model trained on the original data.

On the other hand, the SyntheticDataVault seems to produce data with larger differences to the original, which usually leads to reduced effectiveness. The same is true for the DataSynthesizer when Differential Privacy is enabled. These trends also manifest in direct comparisons of the datasets' properties, e.g., in the heatmaps of pairwise correlations shown in Figure 1.

A basic assumption is that privacy is endangered if the artificial rows in synthetic data are very close or equal to the rows of actual individuals in the original data. Privacy risks could therefore by assessed by computing the distance between each synthetic sample and the most similar original record. Visualisations of these minimal distances can be seen in Figure 2 (the x-axis shows the distance, the y-axis counts the number of records). While the DataSynthesizer without Differential Privacy leads to many records with small distances to original samples, the SyntheticDataVault generates much larger differences.

We complemented this privacy analysis on synthetic data by establishing a baseline for attribute disclosure risks [3]. Attribute disclosure happens when an attacker knows the values of quasi-identifying attributes of their victim (such as birth date, gender or ZIP), and is able to use some data source to infer the value of sensitive attributes (such as personal health data). By considering several scenarios on benchmark datasets, we demonstrated how an attacker might use synthetic datasets for the prediction of sensitive attributes. The attacker's predictive accuracy was usually better for the DataSynthesizer without Differential Privacy and for synthpop than it was for the SyntheticDataVault. However, both the amount of near-matches in the analysis of Figure 2 and the computed attribute disclosure scores show that the risk of reidentification on synthetic data is reduced.

Our evaluations demonstrate that the utility of synthetic data may be kept at a high level and that this approach is appropriate for privacy-preserving data publishing. However, it is important to note that there is a trade-off between the level of the utility and the privacy these tools achieve. If privacy is the main concern, we recommend that samples are generated based on models that preserve fewer correlations. This reduces the attribute disclosure risk and ensures that the artificial records are not too similar to the originals.

Links:

[L1] https://github.com/sdv-dev/SDV
[L2] https://github.com/DataResponsibly/DataSynthesizer
[L3] https://cran.r-project.org/web/packages/synthpop/
[L4] http://archive.ics.uci.edu/ml/datasets/Adult

References:

[1] On the utility of synthetic data: An empirical evaluation on machine learning tasks, ARES '19 Proc., Canterbury, UK, https://doi.org/10.1145/3339252.3339281

[2] Utility and privacy assessments of synthetic data for regression tasks, IEEE BigData '19 Proc., Los Angeles, CA, USA, https://doi.org/10.1109/BigData47090.2019.9005476

[3] A baseline for attribute disclosure risk in synthetic data, CODASPY '20 Proc., New Orleans, LA, USA, https://doi.org/10.1145/3374664.3375722

Please contact:

Markus Hittmeir, Andreas Ekelhart, Rudolf Mayer
SBA Research, Austria
mhittmeir@sba-research.org, aekelhart@sba-research.org, rmayer@sba-research.org

Dronesurance: Using AI Models to Predict Drone Damage for Insurance Policies

by Andrea Fontanari and Kees Oosterlee (CWI)

The increasing use of drones is likely to result in growing demand for insurance policies to hedge against damage to the drones themselves or to third parties. The current lack of accident data, however, makes it difficult for insurance companies to develop models. We provide a simple yet flexible model using an archetype of Bayesian neural networks, known as Bayesian generalised linear models, to predict the risk of drone accidents and the claim size.

Drones (formally known as unmanned aircraft vehicles, Figure 1), are no longer purely recreational objects, but essential business tools, used for purposes from real estate photography to transportation of goods.

The market demand for drones is expected to rise [1], and with increased drone use, comes a concomitant demand for insurance of flights. Additionally, the legislator - the European Union Aviation Safety Agency (EASA) - is advocating a per-flight insurance to cover damage that a drone may cause to a third party. For example, as of 1 July 2020, the guidelines issued by EASA stipulate a per-flight insurance any time a drone is used beyond the line of sight [L1].

The goal of our project, together with our partners Bright Cape (Netherlands), Achmea (Nederlands), University Politecnica, Madrid (Spain) and EURAPCO (Switzerland), is to build a simple but robust prediction model to study the risk embedded in drone flights to help design a per-flight drone insurance.

Insurance pricing is usually determined by a model based on a large pool of historical accident data, using the size and the counts of claims as a function (often non-linear) of the determining features. In the car insurance industry, for example, features may include the type of car, its age and its location. However, with drones being a recent innovation, insurance companies lack adequate data to design a full data-driven model. Data will likely arrive sequentially as more policies are developed and more flights occur. A model therefore needs to be able to handle existing information and expert judgements to partially make up for the initial lack of data, and should adapt to new information as it becomes available.

Figure 1: An unmanned aircraft vehicle (drone) in flight. (Licence Creative Commons made by Walter Baxter).

To this end, we have adopted a flexible framework offered by Bayesian generalised linear models (BGLM) [2]. These models can be understood as simple Bayesian neural networks, with no hidden layers, that can design non-linear interactions between the features and the responses, while maintaining a fair degree of mathematical tractability and explainability.

We chose a sigmoid output layer with a binomial loss function to model the risk of an accident (understood as a probability), and an exponential output layer together with a gamma loss function to model the claim size.

The simple structure of BGLM reduces the risk of over-parametrisation and over-fitting, especially in the initial phases of the training when we cannot expect to have a large batch of available data.

The Bayesian side of the model allows prior information and expert judgements to be incorporated in a simple and natural way: the model parameters are understood as random variables whose distributions, called priors, handle the uncertainty about their realisations.

In terms of the model calibration, the Bayesian framework regularises the loss function to penalise calibration iterations

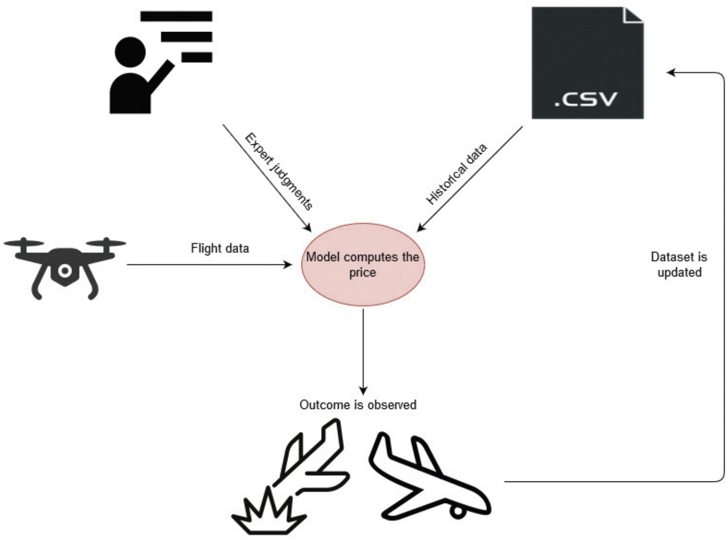


Figure 2: A schematic view of the model.

that tend to over-fit. The Bayesian setting also enables us to handle the sequential nature of the problem. In fact, once the model is calibrated the posterior distribution can be used as a new and improved guess on the parameters for the next run of the model.

Using interviews and questionnaires, we collected expert opinions in order to identify the features that are most likely relevant for our problem. We embedded this information into our model using the hyper-parameters of the prior distributions.

Once data are available, the model is calibrated using a Markov Chain Monte Carlo algorithm (MCMC) and scenarios are generated to predict the risk profile of a drone flight and the size of possible claims that could be generated.

In the early stages, it is likely that the model results will closely resemble the initial guesses formulated by the experts. However, over time, as more data become available, the model will be able to deviate from the initial guesses towards a data-driven explainable estimation by giving more and more weight to the empirical evidence. Finally, Figure 2 exhibits a stylised version of the main ingredients and their interactions of our approach.

While it is important to keep in mind that "all models are wrong but some models are actually useful", it may be reasonable to expect that the model described may not be able to learn very complex nonlinear structures that may arise in the data. An interesting extension that may help to mitigate this issue is the Combined Actuarial Neural Network approach (CANN) [3]. This approach allows a fairly simple model to be blended with a deep feed-forward neural network designed to handle more complex data structures. However, the price to pay for this flexibility is an increase in complexity and the amount of data required for training and a loss of explainability.

Link:
[L1] https://kwz.me/h0h

References:

[1] C. Anderson: "The drone economy. HARv. Bus. REV, 5, 2017.
[2] E. Ohlsson, E., B. Johansson: "Non-life insurance pricing with generalized linear models (Vol. 2)"; Springer, 2010.
[3] M. V.Wüthrich, M. Merz, M.: "Yes, we CANN!", ASTIN Bulletin: The Journal of the IAA, 49(1), 1-3, 2019.

Please contact:

Andrea Fontanari, CWI, The Netherlands
a.fontanari@cwi.nl

DECEPT: Detecting Cyber-Physical Attacks using Machine Learning on Log Data

by Florian Skopik, Markus Wurzenberger, and Max Landauer (AIT Austrian Institute of Technology)

Most current security solutions are tailored to protect against a narrow set of security threats and can only be applied to a specific application domain. However, even very different domains share commonalities, indicating that a generally applicable solution, to achieve advanced protection, should be possible. In fact, enterprise IT, facility management, smart manufacturing, energy grids, industrial IoT, fintech, and other domains, operate interconnected systems, which follow predefined processes and are employed according to specific usage policies. The events generated by the systems governed by these processes are usually recorded for maintenance, accountability, or auditing purposes. Such records contain valuable information that can be leveraged to detect any inconsistencies or deviations in the processes, and indicate anomalies potentially caused by attacks, misconfigurations or component failures. However, syntax, semantics, frequency, information entropy and level of detail of these data records vary dramatically and there is no uniform solution yet that understands all the different dialects and is able to perform reliable anomaly detection on top of these data records.

Today's advanced process security and protection mechanisms for IT systems apply white-listing approaches based on anomaly detection that observe events within a system and automatically establish a baseline of normal user- and system behaviour. Every deviation from this normal behaviour triggers an alert. While there exist numerous behaviour-based anomaly detection approaches for IT security in research [1], they are not easily applicable to other non-IT-centric domains. The reason for this is that these anomaly detection approaches for IT security are usually highly optimised for very specific application areas, i.e. different approaches exist for CPS, cloud security, etc., but they are not adaptive enough to be generally applicable to other domains. Most of them require detailed expertise in the application area and are costly to set up and maintain. Furthermore, most of them analyse network-traffic only, which relies on investigation of domain-specific protocols and becomes ineffective due to the wide adoption of end-to-end encryption. This makes it impossible to track the real system behaviour by inspecting network traffic only. Thus, generally applicable anomaly detection solutions that utilise unstructured textual event logs, created directly by the entities in an environment (e.g., host, camera, control panel etc.) are a promising means to security.

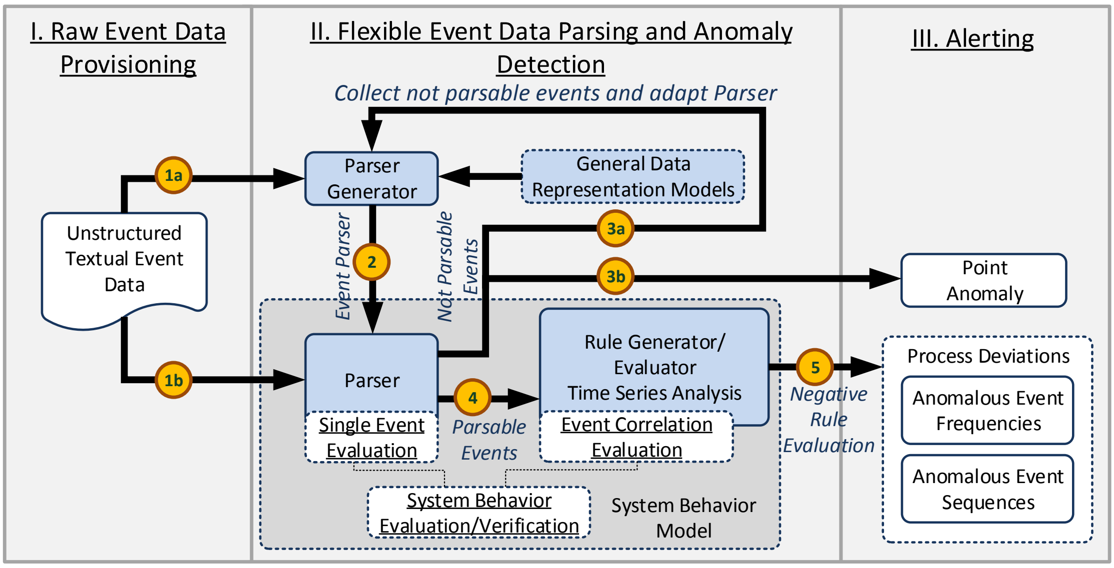


Figure 1: The DECEPT methodology.

DECEPT technical objective

The overall goal of DECEPT is to develop a generally applicable concept of an anomaly detection (AD) approach that can be applied to various domains, and to implement a proof of concept that demonstrates the ability of the DECEPT approach to analyse and evaluate work processes, as well as to monitor environmental events, in different application areas. Figure 1 illustrates the methodology to achieve this objective. The DECEPT approach will analyse unstructured textual event data, such as syslog messages from computer systems or protocol data from manually recorded events (e.g., access logs). In the training phase (1a) a parser generator [2] analyses the text data and automatically builds event parsers, i.e., identifies implicit structures in apparently unstructured records. Then, it uses general data representation models as building blocks to iteratively increase the comprehension of text structures and embedded data types (such as dates, times, identifiers etc.). This way, parsers can be created to decompose and understand textual representations of events with no manual intervention.

Applying parsers generated in this way to different sources of unstructured textual event data, allows data of different types, sources and domains to be correlated. After the training phase (2), the parser obtains the event parsers from the parser generator. The parser then analyses the unstructured textual event data entities separately (1b), i.e. it performs a single event evaluation. Thus, depending on the configuration, the parser either forwards not parse-able events to the parser generator, which collects them and adapts the event parser (3a), or it triggers a point anomaly (3b). The rule generator/evaluator and time series analysis module [3] obtains parse-able events from the parser and defines and evaluates statistical rules. For example, it evaluates the distribution with which events occur, it defines correlation rules (e.g., timely correlation of variable parts of the event data), and it carries out a time series analysis. Thus, the module performs event correlation evaluation, allowing detection of deviations of complex processes from the normal system behaviour; these deviations manifest in anomalous event frequencies and sequences (5). Eventually, the parser along with the rule evaluator module, define a model of the normal behaviour of the observed environment (e.g., a computer network, or a facility), evaluate the model and continuously verify it.

DECEPT will demonstrate in the course of a proof of concept the general applicability of the anomaly detection approach in two independent application areas: (i) Enterprise IT security and (ii) (IT-supported) facility security. Since modern attacks often exploit vulnerabilities in different application areas, processes of, for example, IT security and facility security must be aligned to allow a timely reaction to potential attacks. Some examples of such attacks are the remote manipulation of IP-protocol based access control systems to aid physical intrusion, or the physical access to and manipulation of switches to make them vulnerable to cyber-attacks.

The project DECEPT and its consortium

In order to attain these ambitious goals and finally ensure the wide applicability of developed tools and procedures, the project consortium consists of a vital mix of a strong academic partner with deep knowledge in cyber security and machine learning (Austrian Institute of Technology), an enterprise security solution vendor (Huemer iT-Solutions) and a vendor of physical security equipment (PKE Holding AG). DECEPT is a 30-month national research project running from 2020 to 2022 and is funded by the Austrian FFG Research Program “ICT of the Future”.

References:

[1] R. Mitchell and I.R.Chen: “A survey of intrusion detection techniques for cyberphysical systems”, ACM Comp. Surv. 46, 4, Article 55, 2014.

[2] M. Wurzenberger et al.: “AECID-PG: A Tree-Based Log Parser Generator to Enable Log Analysis”, in 2019 IFIP/IEEE Symposium on Integrated Network and Service Management (IM) (pp. 7-12), IEEE.

[3] M. Landauer et al.: “Dynamic log file analysis: An unsupervised cluster evolution approach for anomaly detection”, computers & security, 7(9), 94-116, 2018.

Please contact:
Florian Skopik
AIT Austrian Institute of Technology, Austria
+43 664 8251495, florian.skopik@ait.ac.at

Photographic Heritage Restoration Through Deep Neural Networks

by Michaël Tits, Mohamed Boukhebouze and Christophe Ponsard (CETIC)

Recent advances in deep neural networks have enabled great improvements in image restoration, a long-standing problem in image processing. Our research centre has experimented with cutting edge algorithms to address denoising, moire removal, colourisation and super resolution, to restore key images representing the photographic heritage of some of Belgium's top athletes.

The research field of image restoration is benefiting from recent major improvements in artificial intelligence algorithms, particularly deep learning, a family of black-box algorithms using deep neural networks allowing a machine to learn relevant relationships and abstractions, provided enough examples are available.

Such algorithms can improve poor quality images by learning how to "invent" detail from the initial image, for instance by retrieving colours of a monochrome image, or filling in missing pixels realistically. These techniques are still being developed, but they show great promise in different contexts that require improved image quality, such as video surveillance, medical and satellite imagery and culture (history and art), which was our experimentation field.

As part of the DigiMIR project [L1], CETIC (ICT R&D centre) and Numediart (multimedia multimedia research institute) have collaborated to select, test and compare recent techniques in this field, with the aim of developing a generic and innovative image restoration tool going beyond prototypes proposed in the scientific literature. Tools to date have generally been confined to a very specific algorithm (e.g. [1]) or prototype (e.g. [L2]) and are not integrated in a consistent processing chain. A more generic approach can have advantages, for example noise reduction can improve the texture of different surfaces, such as that of a lawn, which will also improve the results of a colourisation algorithm. The more comprehensible the image, the better the algorithm will work! Conversely, an image that is too noisy, or too smoothed by a poorly executed "denoising" algorithm, will prevent the lawn texture from being recognised, resulting in greyish colours.

Our work focused on the cascading of four types of techniques to address the following defects: unstructured noise (grain/ white noise), structured noise (stripes), monochromatic character and low resolution. For each aspect, a comparative analysis was carried out to identify the most relevant algorithms. We then developed a pipeline allowing an end-to-end restoration of the images and making it easy to cascade techniques, and to couple them with more conventional image processing. This active project is available as open source [L3].

A concrete application case was carried out in collaboration with a Belgian cultural organisation and Jean Vandendries, the author of a book detailing the careers of athletes from the past century [2]. The book was facing the problem of printer rejection of about half of the images (i.e., 50) due to poor quality. The images contained a wide variety of defects, including pronounced grains, several scratches and streaks due to the aging of a physical medium, numerous monochrome images and very low resolution. As an example, Figure 1 shows different results of stripe removal from the original image on the left. This study was rewarding because we could successfully restore historical material. In addition, this challenging material was a strong driver for building our toolchain and considering many practical issues, such as avoiding the injection of new artefact and automatically assessing the quality of the restoration.

Links:
[L1] https://www.cetic.be/DigiMIR-2789
[L2] https://github.com/jantic/DeOldify
[L3] http://www.mickaeltits.be/open-image-restoration

References:
[1] J. Guan, R. Lai and A. Xiong: "Wavelet Deep Neural Network for Stripe Noise Removal," in IEEE Access, vol. 7, pp. 44544-44554, 2019.
[2] J. Vandendries: "Les Étoiles sportives nivelloises", 2019.

Please contact:
Christophe Ponsard, CETIC, Belgium
+32 472 56 90 99
christophe.ponsard@cetic.be

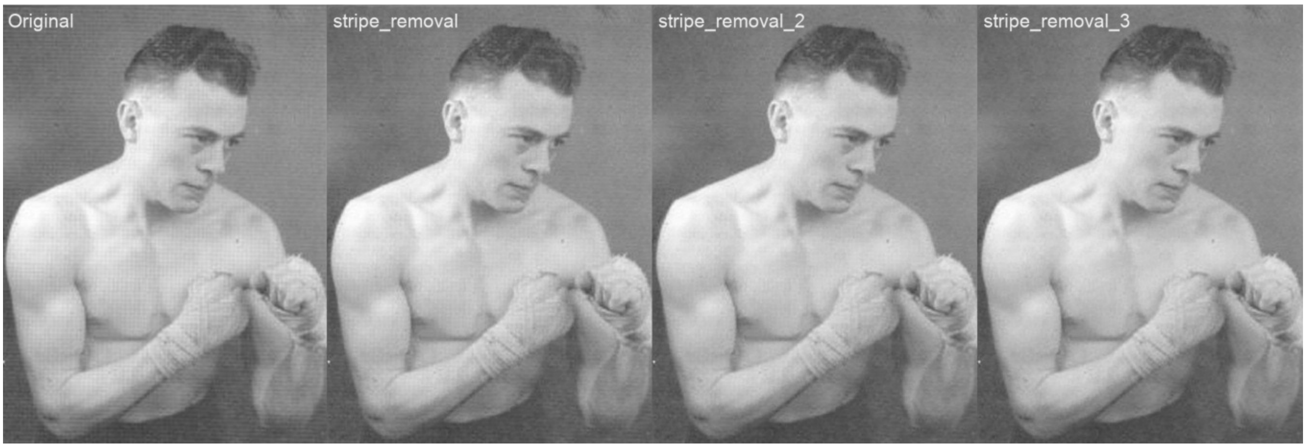


Figure 1: Example of stripe removal (Paul Goffaux, Belgan boxing champion in 1942).

Penetration Testing Artificial Intelligence

by Simon Tjoa (St. Pölten UAS, Austria), Christina Buttinger (Austrian Armed Forces), Katharina Holzinger (Austrian Armed Forces) and Peter Kieseberg (St. Pölten UAS, Austria)

Securing complex systems is an important challenge, especially in critical systems. Artificial intelligence (AI), which is increasingly used in critical domains such as medical diagnosis, requires special treatment owing to the difficulties associated with explaining AI decisions. Currently, to perform an intensive security evaluation of systems that rely on AI, testers need to resort to black-box (penetration) testing.

In recent years, artificial intelligence (AI) has significantly changed the way we do business and research. Applications that previously seemed possible only in science fiction (e.g. personal assistants like Siri and Alexa) are now a reality. AI components are also becoming increasingly important in the automated decision-making routines behind many systems that are used in areas such as cancer research, open-source intelligence (OSINT) and intrusion detection.

However, there is one huge drawback that limits the use of this technology. Often it remains unclear what exactly deep neural networks or similar approaches have learned and whether the software can be trusted. For some applications, either substantial research is conducted to gain a deeper understanding of their inner workings, or a human is involved in the process to ensure valid operation (i.e. 'human-in-the-loop'). While this approach is feasible in many cases, e.g. the doctor-in-the-loop, many applications, especially those that concern decision-making in critical infrastructures, do not scale with a human in the loop, often due to their time-critical nature. Furthermore, many of these decision-making processes need to be based on large amounts of inferenced datasets, thus making manual analysis practically impossible. This greatly reduces the trust in the results derived from such systems. In addition, in some applications, such as self-driving vehicles, it is not possible to use explainable AI or human intervention. Therefore, it is crucial to use an attacker's mindset to test the robustness and trustworthiness of the artificial system – especially considering the large attack surface posed by these systems and the massive developments in adversarial machine learning [1]. Combined with the inability to explain results, a lot of damage could be caused by attackers manipulating intelligent systems for their own gain.

We propose a high-level concept of a systematic process to test AI systems within the data science lifecycle. This is mainly done by combining techniques from risk management (i.e. assessing the business risk, existing controls and the business case for an attacker), adversarial machine learning (i.e. evaluating the trustworthiness and robustness of the algorithm and trained abilities) and traditional penetration testing (i.e. evaluating the security of the implemented system, e.g. manipulation of sensor data).

Figure 1 gives an overview of the generic approach, focussing on the AI components. While standard penetration testing of the underlying platform is required to mitigate threats and security gaps on this level (the greyed-out part labelled 'platform' in Figure 1), this method extends the standard approaches to achieve certain tasks required for the AI components. The main problem with AI components is explainability; it is usually not possible to gain a detailed understanding of why a certain decision was made [2]. Thus, testers resort to black-box security testing, trying to generate unwanted results either by using completely random (fuzzied) input material or by using nearby or extreme values. When using algorithms that learn from past decisions, it is vitally important to attack the underlying knowledge. We must assume that an attacker might be in possession of parts of the underlying (training) data or even have a (black-box) environment running the algorithms in question. The latter would enable the attacker to run many executions using arbitrary data or fuzzied information, trying differential attacks and feeding specially structured information into

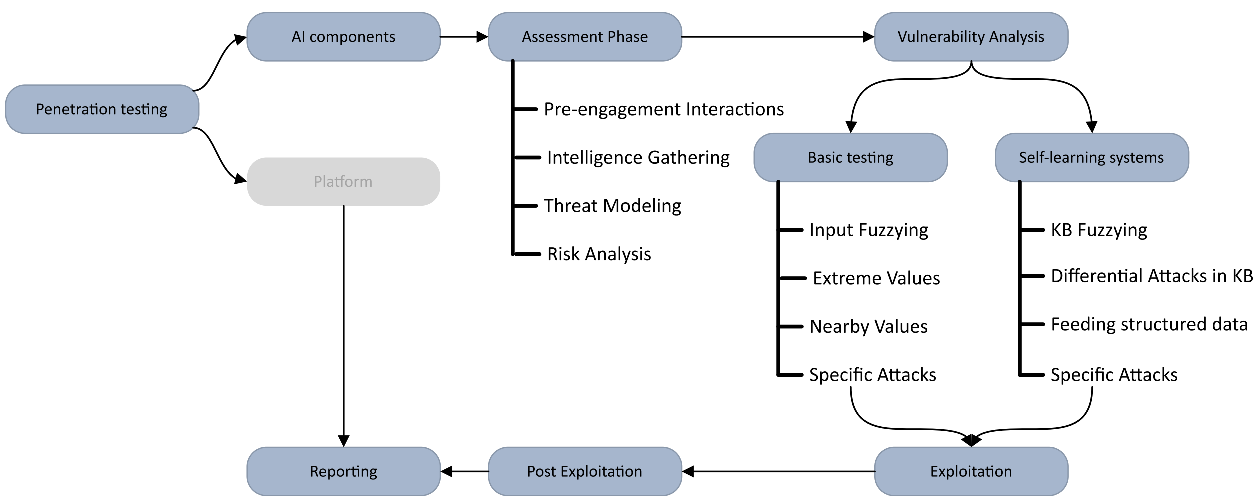


Figure 1: A high-level approach to penetration testing AI systems.

the system. This is very similar to the cryptoanalytic counterparts of partially known and chosen plaintext attacks. Furthermore, depending on the algorithms in use, specific attacks might exist that need to be considered during the penetration test.

While penetration testing is extremely valuable to evaluate such systems, proper risk analyses are often overlooked. These are important to: (i) carve out the attack surface, and (ii) help determine mitigation strategies and possible attack scenarios. Further research into possible attack scenarios is particularly important as the potential damage caused by manipulation of intelligent systems is often not clear even for the system's designers. Possible outcomes range from the introduction of broad bias into decision-making processes through to an attacker being able to launch fine-tuned attacks. Thus, together with identifying the (information) assets, the security analyst will also need to determine possible attack and damage scenarios in order to develop a feasible mitigation strategy.

The proposed workflow is at first draft stage and requires additional methods to tailor it to specific systems and the technologies. Nevertheless, it can be used as a template to provide a basic level of security in AI-based systems. Importantly, penetration testing can never give a security guarantee; at best, the testers will find all bugs that could have been found by attackers, but as history has shown, even very prominent software stacks can be susceptible to newfound or newly introduced errors [3]. We are investigating these issues in two academic projects, the COIN-project "Big-Data Analytics" [L1] and the FORTE-project "exploreAI" [L2]. In these projects we are conducting in-depth research into efficient penetration testing against intelligent systems and future implications for critical infrastructure security.

Links:
[L1] https://research.fhstp.ac.at/en/projects/big-data-analytics
[L2] https://kwz.me/h1Q

References:

[1] A. Chakraborty et al.: "Adversarial attacks and defences: A survey", arXiv preprint arXiv:1810.00069, 2018.
[2] K. Holzinger et al.: "Can we trust machine learning results? artificial intelligence in safety-critical decision support", ERCIM NEWS, (112), pp.42-43, 2018.
[3] Z. Durumeric et al.: "The matter of heartbleed", in proc. of the 2014 conference on internet measurement (pp. 475-488), 2014.

Please contact:

Simon Tjoa
University of Applied Sciences St. Pölten, Austria
simon.tjoa@fhstp.ac.at

Future Cyber-security Demands in Modern Agriculture

by Erwin Kristen, Reinhard Kloibhofer (AIT Austrian Institute of Technology, Vienna) and Vicente Hernández Díaz (Universidad Politécnica de Madrid)

The European agricultural sector is transforming from traditional, human labour-intensive work to data-oriented digital agriculture that has great potential for semi- or fully autonomous operation. This digital transformation offers many advantages, such as more precise fact-based decision making, optimised use of resources and big changes in organisation – but it also requires improved cyber-security and privacy data protection.

To feed the world's growing population and compensate for the loss of arable soil, the agricultural sector needs to increase efficiency, productivity and food quality, while simultaneously reducing labour costs and environmental impacts. The current approaches aim to use more powerful machines in the field, make these machines semi- or fully autonomous, and to plan precise fertilisation, irrigation, pest control and harvesting regimes based on detailed environmental data. The race for solutions has started: fields, crops and livestock are supplied with numerous sensors that monitor the environment. Machines are equipped with intelligent algorithms to perform their daily work with high precision and provide extensive operation status information, enabling a 24/7 availability. The agricultural system infrastructure, composed of numerous networked digital devices, is called Agriculture Internet of Things (AIoT).

The resulting bulk data can guide precise decision making on the farm and inform product development by machinery manufacturers. However, the colossal data gathering activities are very attractive for cyber-attacks, including theft, manipulation and misuse of data.

In the "AFarCloud" project, a group of European partners are working to implement the AIoT concept. We are currently developing an abstraction layer for AIoT-based architectures. This is the middleware that defines the software components and procedures, acting as an interface between the field layer and the cloud-based data processing layer where the farm management services are located (Figure 1).

The field layer includes the sensors, actuators, outdoor devices, vehicles and livestock. One important part of the middleware is the cross-layer cyber-security management (CSM) service, which handles the security maintenance process, providing a security process definition for periodic security assessment and security improvement recommendations. It facilitates a trouble-free, secure operation.

Cyber-security measures protect the production plant against attacks. In the early days of automation, only the information technology (IT) sector (farm management and middleware) was affected by cyber-security threats. The operational technology (OT) domain (farm and field) was secure by virtue of isolation. Today, cyber-security measures are necessary for both domains.

Modern sensors are more than just data sources that send data to a data gateway: they are small, complex systems with a micro-controller, FLASH and RAM memory, that hold an operating system and data pre-processing firmware. They need the capacity to update their onboard firmware. Because of the increasing number of sensors in the field, this must be done using a wireless method, such as "over the air" (OTA). Unfortunately, this contributes considerably to security risks.

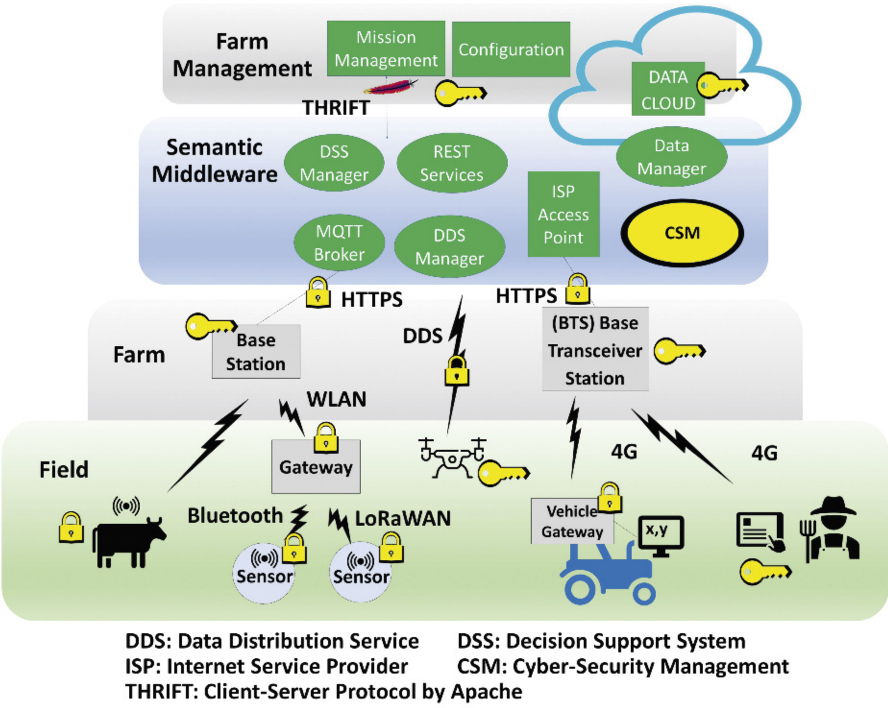


Figure 1: The system architecture with the middleware as the interface between the field level and the cloud-based farm management.

Cyber-security risks in general are not addressed by the European agricultural standards, which focus on food and nutrition security, prevention of harm to workers as a result of farm labour or exposure to pesticides, minimising the use of heavy machines and ensuring the humane treatment of livestock. In the United States, the Department of Homeland Security (DHS) recently conducted research to identify potential cyber-security vulnerabilities for agriculture. In Europe, however, agriculture was absent from a recent paper [1] that outlined the risks and the need for monitoring support to ensure cyber-security for a range of domains. Even the EU publication "Study on risk management in EU agriculture" [2] failed to include smart farming or cyber-security.

There is a clear need to define cyber-security guidelines for modern Agriculture 4.0 in the EU. While dedicated IT/OT security standards are lacking in the agricultural sector, industrial automation control systems (IACS) and the automotive domain are guided by cyber-security standards to ensure secure operation. These could be partially transferred to agricultural electronic systems by manufacturers. Similarly, in the communication sector there exist ETSI M2M (machine-to-machine) communication standards with extensions for agricultural machines [L2] and the ISO Bus.

The AFarCloud project has been running for two years. In the third and final year we will focus on a security evaluation demonstrator (SED) [3] that shows how installing a fast-responding sensor manipulation monitoring system can improve the security of hardware and software in a simple sensor node.

We also plan to begin developing an agriculture cyber-security standard and to publish our research results (cyber-security assessment and analysis methodologies, requirements, security recommendations) as a useful guide for cyber-security in agriculture.

This project has received funding from the ECSEL JU (Horizon 2020) under grant agreement No 783221 and from the partners' national funding organisations.

Links:
[L1] http://www.afarcloud.eu/
[L2] ETSI TR 103 511, V1.1.1 (2018) - SmartM2M - SAREF extension investigation - Requirements for AgriFood domain, https://kwz.me/h2k

References:

[1] I. Nai-Fovino et al.: "European Cybersecurity Centres of Expertise Map - Definitions and Taxonomy", doi:10.2760/622400. https://kwz.me/h1R
[2] Directorate-General for Agriculture and Rural Development (EC, 2017), ECORYS, Wageningen Economic Research, Study on risk management in EU agriculture, doi:10.2762/08778. https://kwz.me/h1U
[3] R. Kloibhofer, E. Kristen, L. Davoli: "LoRaWAN with HSM as a Security Improvement for Agriculture Applications", 2020. http://doi.org/10.5281/zenodo.3999637

Please contact:

Erwin Kristen
AIT Austrian Institute of Technology
erwin.kristen@ait.ac.at

Vicente Hernández Díaz
Universidad Politécnica de Madrid
vicente.hernandez@upm.es

FLEXPROD - Flexible Optimizations of Production with Secure Auctions

by Thomas Lorünser (AIT), Niki Dürk (X-Net), Stephan Puxkandl (Ebner)

The FlexProd research project aims to develop a platform to improve the efficiency and speed of cross-company order allocation and processing in the manufacturing industry. The platform will help make production more flexible, facilitating cooperation between manufacturers and customers to enable a more agile and efficient manufacturing industry.

To serve its clients, the manufacturing industry must comply with production standards and defined delivery and production times. To ensure that production peaks and failures of individual plants can be managed, factories are designed with redundancies and parallelly equipped production lines. As a result, individual plants have a surplus of production capacity: in some cases up to 50 or 60% of idle capacity per year. At the same time, due to increasing digitalisation and cross-company value creation networks, demands on the manufacturing industry are increasing.

Improved internal and external communication can lead to more flexible, individual and efficient production of smaller batches – and ultimately a more competitive and sustainable economy. Processes and procedures for the simple, flexible and rapid placement and processing of orders and outsourcing are required in order to respond quickly to fluctuations in production capacity – bottlenecks or underutilisation – and to increase the overall efficiency of existing facilities.

FlexProd is a platform that is being developed to interface connecting clients and plant operators. The system will ensure fast processing of (supra-regional) projects and enable more efficient use of resources. Its design will be based on strong technical methods, strictly following security by design approaches [1].

The research and development of FlexProd is based on the following design paradigms:

- Intermediation: The rapid and (fully) automatic processing of orders is intended to reduce underutilised facilities, avoid vacancies and increase overall efficiency.
- Speed: Fast support for the duration of the contract means savings in time and money during sales transactions.
- Security: The producer should retain control of their data at all times; it must not leave the company in an unsecured form. The platform should also prevent competition analysis (e.g. machine types/equipment, capacities, prices).
- Transparency: Secure order processing should be achieved through diversity of participants and previously known processing costs and methods. By disclosing the methods, processes, interfaces and software implementations, trust in the platform should be created.
- Additional service and benefits: The integration of additional services, such as insurance brokerage, financing, quality assurance and logistics, should provide the customer with additional benefits.

The platform architecture decentrally identifies the appropriate production plant and the matching service offer for orders to be awarded (MatchMaking). The industrial sector currently lacks comprehensive profiles for the representation of heterogeneous assets in terms of matchmaking technological and economic conditions. Formalisms for the interorganisational representation of MatchMaking assets on different levels of abstraction and multiple frameworks are needed. Additionally, there is a lack of decentralised recommendation mechanisms that allow an interorganisational coordination of outsourcing supply and demand. This task requires a hybrid MatchMaking solution with semantic similarity measures and dynamic trust mechanisms for binding matching of industrial assets.

The platform supports customers and suppliers through the actual establishment of a transaction, without disclosing data

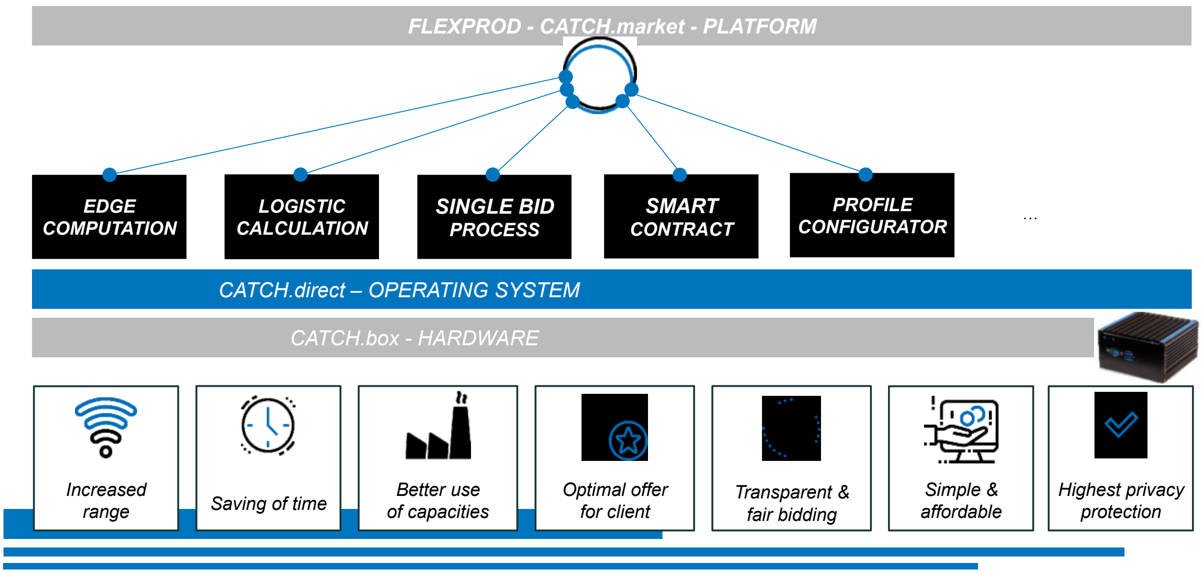


Figure 1: Features of CATCH.direct platform as developed in FlexProd.

inadvertently or divulging trade secrets (by providing a secure anonymous auction service). This is achieved through multiparty computation (MPC), which protects sensitive information while still participating in a common market place, leveraging an auction mechanism to determine the best market prices. However, because MPC technology does not scale well and is rather slow, carefully designed systems and protocols are required to handle larger marketplaces [2]. Furthermore, MPC can be extended using methods from verifiable computing, which improves the traceability for individual transactions without sacrificing the privacy of the data.

The integration of blockchain improves the traceability and transparency for trust and acceptance of the platform. The project's approach is innovative in terms of its high degree of decentralisation and the development of security algorithms and standards to best protect sensitive business data (e.g., corporate data, production data or project calculations). The decentralised approach also makes it harder to assure the correct working of the system and provide auditing and verification means for market participants. Using a blockchain to log and trace transactions can increase the trustworthiness; however, deployment is an important issue to consider [3]. Nevertheless, it can also serve as a root of trust for the management of identities. It will be used for identity and key management to enable an open platform. Current developments, such as decentralised identifiers (DIDs), will be considered.

FlexProd is a joint effort between industry and academia and is being developed in consultation with all relevant stakeholders. It will contribute to a more agile and efficient manufacturing industry and lead to more efficient resource use, thus also contributing to the European Green Deal.

Links:
[L1] https://www.flexprod.at
[L2] https://catch.direct

References:
[1] P. Chiaro, et al: "Secure and privacy-friendly storage and data processing in the cloud", IFIP Advances in Information and Communication Technology, 526, 153–169, 2018. https://doi.org/10.1007/978-3-319-92925-5_10
[2] T. Lorünser, F. Wohner: "Performance Comparison of Two Generic MPC-frameworks with Symmetric Ciphers", in proc. 17th Int. Joint Conf. on E-Business and Telecommunications, 587–594. https://doi.org/10.5220/0009831705870594
[3] E. Framner, et al.:"Making secret sharing based cloud storage usable", Information & Computer Security, 27(5), 647–667, 2019. https://doi.org/10.1108/ICS-01-2019-0016

Please contact:
Thomas Lorünser
AIT Austrian Institute of Technology GmbH
Thomas.Loruenser@ait.ac.at

HEALPS 2: Tourism Based on Natural Health Resources for the Development of Alpine Regions

by Daniele Spoladore, Elena Pessot and Marco Sacco (STIIMA-CNR)

The HEALPS 2 project is using digital solutions and stakeholder engagement to unlock the potential of health tourism in the alpine regions.

Health tourism is growing exponentially worldwide [1], as a consequence of changing lifestyles, the desire for nature-based experiences and the aging population. The wider definition of health tourism includes wellness tourism – dedicated to maintaining or enhancing personal health and well-being – and spa tourism – which exploits spa waters to heal or relax with the aim of curing or preventing health problems. The alpine regions of Europe, characterised by blue spaces (e.g. streams, pounds, lakes, etc.), high altitudes, and pristine landscapes, have the potential to play a pivotal role in nature-based health tourism. Nevertheless, the spatial fragmentation, the lack of access to knowledge and the little transversal cooperation between different stakeholders are hampering development of these regions.

The European project HEALPS 2 aims to help this industry grow by providing innovative ways of developing health tourism products and service chains. We are creating up-to-date digital solutions, informed by key stakeholders and recent research in the field of alpine health tourism. The outcome will be a "tactical health management toolset" to unlock the potential of alpine assets for value generation and best practice sharing. The digital tools will be developed and integrated in close collaboration with both existing and potential stakeholders, including tourism associations, policy-makers, regional developers, sectoral agencies and SMEs. The proposed approach will be tested in pilot regions where tourism, health and other relevant sectors will be engaged in a cross-fertilisation positioning of the alpine space as a globally recognised health-promoting region.

The set-up work of HEALPS 2 deals with the assessment of the core nature-based resources of alpine regions, including forests, moderate and high altitude , blue spaces (lakes, rivers, streams, ponds and waterfalls). An effective use of these resources can have positive impacts on visitors' general health and can improve specific physical problems, such as back and joint pain, overweight, low levels of fitness, allergies and asthma and recovery following leg injury [2]. HEALPS 2 will help the development of nature-based health tourism products and fill in the knowledge gaps around marketing of health tourism in these regions by describing a set of KPIs to frame the nature-based regional features, together with the health-related services already provided by the SMEs operating in the regions (such as hydrotherapy, phys-

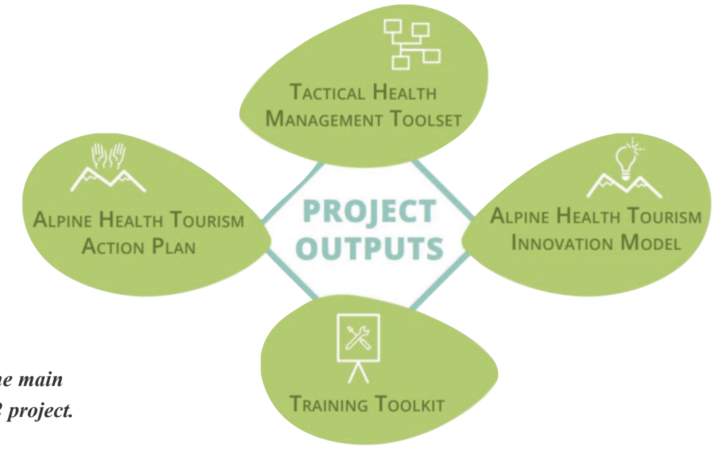


Figure 1: An overview of the main outcomes of the HEALPS 2 project.

iotherapy, healing caves and nature-based products). This information, combined with data about the benefits the services provide to specific target groups (tourists looking for nature-based health tourism to cure physical ailments), is invaluable when it comes to identifying development opportunities for alpine health tourism.

This knowledge is modelled and stored in a semantic data model that leverages the semantic web's capabilities to formalise information (with first-order logic based languages RDF and OWL) and to enable automatic reasoning processes, including health-related data [3]. This feature can be used to provide SMEs and alpine tourist locations with a set of guidelines and reference practices on how to exploit the natural potential of the regions. Similarly, it is possible to advise health tourism operators about which target group could benefit most from the natural assets and services at a specific location, or which target groups could benefit from the collaboration between stakeholders at different levels. Moreover, data analytics techniques can allow SMEs and regional tourism agencies to generate sound what-if scenarios to match locations' services with tourists' demand. Specific factors and possible (often unknown) drivers for further development of the health tourism sector in the alpine regions can be also extracted. These outputs will also provide a sound base of knowledge for policy makers to address supporting measures and funding opportunities for innovation in health tourism and related sectors.

The combination of digital tools and systematic stakeholder engagement will contribute to a health tourism innovation model for the alpine space. The model will provide good and best practices, strategies and techniques to enhance the capacity of alpine regions in implementing innovative health tourism value chains with a transnational approach. The needs of protected areas, the perspectives of thermal regions and the challenges facing peripheral rural areas will also be considered. The innovation model will play a fundamental role in the drafting of the Alpine Health Tourism Action Plan. This document will present the strategy for positioning the European Alpine Space as a globally recognised health tourism destination. It represents a key example of innovation for increasing value from nature-based assets to address health-related problems and enhance economic development. Finally, training modules within a training toolkit will address skill gaps in the use of the innovation model, management toolset and Action Plan. The modules will be tailored for stakeholders at several levels and in different regions.

HEALPS 2 is a collaborative project co-financed by the European Union via Interreg Alpine Space. It is led by Paracelsus Medical University (Salzburg, Austria), and involves research centres and digital SMEs (National Research Council of Italy, STIIMA; Scientific Research Centre Bistra Ptuj; University of Applied Sciences HTW Chur; MOXOFF Spa), regional stakeholders' networks (ALPARC – the Network of Alpine Protected Areas; Community Network Alliance in the Alps; Management Body of the Ossola Protected Areas.) and development centres (Innovation and Technology Transfer Salzburg; Development centre Murska Sobota) located in six alpine countries (Austria, Italy, Germany, Slovenia, France, Switzerland).

Link:
https://www.alpine-space.eu/projects/healps-2/en/home

References:
[1] Research for TRAN Committee – Health Tourism in the EU: a general investigation. Available online: https://www.europarl.europa.eu/RegData/etudes/STUD/2017/601985/IPOL_STU(2017)601985_EN.pdf
[2] A. Hartl, et al. "Health effects of alpine waterfalls." Hohe Tauern National Park 5th Symposium, Conference Volume for Research in Protected Areas. 2013.
[3] D. Spoladore: "Ontology-based decision support systems for health data management to support collaboration in ambient assisted living and work reintegration", Working Conference on Virtual Enterprises. Springer, Cham, 2017.

Please contact:
Daniele Spoladore, STIIMA-CNR, Italy
daniele.spoladore@stiima.cnr.it

ERCIM “Alain Bensoussan” Fellowship Programme

The ERCIM PhD Fellowship Programme has been established as one of the premier activities of ERCIM. The programme is open to young researchers from all over the world. It focuses on a broad range of fields in Computer Science and Applied Mathematics.

> “ERCIM Fellowship Programme provided me an amazing experience to work with successful scholars in one of the well-developed universities in Europe. The acquired skills both in technical and non-technical aspects helped me to improve my self-confidence. I strongly recommend this programme to all researchers who are open and willing to learn new skills for developing their career plan.
>
> **Yaser DALVEREN**
> Former ERCIM Fellow

The fellowship scheme also helps young scientists to improve their knowledge of European research structures and networks and to gain more insight into the working conditions of leading European research institutions.

The fellowships are of 12 months duration (with a possible extension), spent in one of the ERCIM member institutes. Fellows can apply for second year in a different institute.

Conditions

Candidates must:

- have obtained a PhD degree during the last eight years (prior to the year of the application deadline) or be in the last year of the thesis work with an outstanding academic record;
- be fluent in English;
- have completed their PhD before starting the grant;
- submit the following required documents: cv, list of publications, two scientific papers in English, contact details of two referees.

The fellows are appointed either by a stipend (an agreement for a research training programme) or a working contract. The type of contract and the monthly allowance/salary depends on the hosting institute.

Application deadlines

Deadlines for applications are currently 30 April and 30 September each year.

Since its inception in 1991, over 500 fellows have passed through the programme. In 2019, 53 young scientists commenced an fellowship and 79 fellows have been hosted during the year. Since 2005, the programme is named in honour of Alain Bensoussan, former president of Inria, one of the three ERCIM founding institutes.

http://fellowship.ercim.eu

> “The incredible ERCIM fellowship gave me the chance to meet the amazing people with whom I will be collaborating for the next few years of my career. Indeed, this is a great opportunity for all talented researchers to pursue their dreams of becoming part of the global research network.
>
> **Fereshteh MIRJALILI**
> Former ERCIM Fellow

SCHLOSS DAGSTUHL
Leibniz-Zentrum für Informatik

Call for Proposals

Dagstuhl Seminars and Perspectives Workshops

Schloss Dagstuhl – Leibniz-Zentrum für Informatik is accepting proposals for scientific seminars/workshops in all areas of computer science, in particular also in connection with other fields.

If accepted the event will be hosted in the seclusion of Dagstuhl’s well known, own, dedicated facilities in Wadern on the western fringe of Germany. Moreover, the Dagstuhl office will assume most of the organisational/administrative work, and the Dagstuhl scientific staff will support the organizers in preparing, running, and documenting the event. Thanks to subsidies the costs are very low for participants.

Dagstuhl events are typically proposed by a group of three to four outstanding researchers of different affiliations. This organizer team should represent a range of research communities and reflect Dagstuhl’s international orientation. More information, in particular, details about event form and setup as well as the proposal form and the proposing process can be found on

https://www.dagstuhl.de/dsproposal

Schloss Dagstuhl – Leibniz-Zentrum für Informatik is funded by the German federal and state government. It pursues a mission of furthering world class research in computer science by facilitating communication and interaction between researchers.

Important Dates

- Proposal submission: April 1 to April 15, 2021
- Notification: July 2021
- Seminar dates: In 2022/2023 (final dates will be announced later).

Special Issue on "Test Automation: Trends, Benefits, and Costs"

Journal of Systems and Software – Elsevier

Today software has a significant impact on all aspects of our society, and its functioning is crucial for a multitude of economic, social, and educational activities. As a consequence, correctness and quality assurance for software systems becomes paramount.

Automation across the Software-Testing Process is a powerful asset to improve the quality of software systems: while originally conceived for test execution, nowadays it is increasingly used for a wide range of testing activities (e.g., generation, prioritization, evolution). However, also the development and the maintenance of test code require considerable effort and skills. Besides, tests themselves need to be kept aligned with the ever-evolving system under test. Therefore, the potential benefits aimed by test automation have to be weighted against its costs and drawbacks.

This special issue aims at investigating some pressing needs for the research in test automation: i) how to integrate means to assess the usefulness and the applicability of new techniques and approaches; ii) what are the issues that prevent broader adoption of existing tools or solutions, and iii) how to collect evidence of (financial, technical and social) costs and benefits of the current trend and practices in test automation. Some of the answers to these questions come from contributions regarding approaches, techniques, tools, experience reports, or empirical studies about test automation adoption and its effectiveness.

Guest Editors
- Maurizio Leotta, University of Genova
- Guglielmo De Angelis, CNR-IASI
- Filippo Ricca, University of Genova
- Antonia Bertolino, CNR-ISTI

Deadline
Manuscript submission deadline:
31 March 2021

Complete call for papers and submission information:
https://kwz.me/h2s

Ton de Kok Appointed New Director of CWI

© Angeline Swinkels

Prof. Ton de Kok has been appointed as director of CWI, by the board of CWI's mother organization NWO-I, the institutes organization of the Netherlands Organisation for Scientific Research. On 1 October 2020, De Kok succeeds current director Prof. Jos Baeten, who has led the institute since 2011 and who will retire.

After finishing his PhD at the Vrije Universiteit Amsterdam, De Kok joined Philips, where he explored ways to better connect logistics and production with each other, as well as with the market. Since 1991, De Kok has been a professor of Quantitative Analysis of Logistical Management Problems at Eindhoven University of Technology.

De Kok says: "As the new director of the CWI, I hope to contribute to the positioning of computer science and mathematics as scientific disciplines that are crucial to understanding the world around us and making it a little more beautiful. By describing laws in the unambiguous language of mathematics, I have learned how we can use scarce resources as efficiently as possible to continue to provide mankind with food, clothing and means of communication".

On the ERCIM Board, Jos Baeten has been succeeded by Prof. Han La Poutré in the beginning of this year, in the position of Vice President.

TRAPEZE - Transparency, Privacy and Security for European Citizens

ERCIM is participating in the EU-funded Innovation Action "TRAPEZE" . The ambitious goal of the project is to drive a cultural shift in the protection of the European data economy by weaving trust into its very foundation and reconstructing the concepts of control, transparency, and compliance through technical and methodological, citizen-first, innovations.

Data is an essential resource for economic growth, competitiveness, innovation, job creation and societal progress in general. Common European data spaces will ensure that more data becomes available for use in the economy and society, while keeping companies and individuals who generate the data in control. TRAPEZE project will develop technologies to empower citizens to actively contribute to the cyber resilience of the common European data space. It will provide the public with the necessary tools and information to manage their security and privacy. It will also create technologies to enforce the integrity and non-repudiation of citizens' data usage policies and processing across data sources. Overall, the project will carve out a path to put cutting-edge technologies like blockchain and linked data to practical use.

Driven by the needs of three distinct real-world use cases, TRAPEZE brings together over a decade worth of EU-funded research in security and privacy, as well as proprietary solutions and know-how, towards realistic and marketable solutions. The project is coordinated by TENFORCE, Belgium. ERCIM is the administrative and finanacial coordinator of the project. The project started in September 2020.

https://trapeze-project.eu/

Please contact:
Jessica Michel Assoumou, ERCIM Office, jessica.michel@ercim.eu

ERCIM – the European Research Consortium for Informatics and Mathematics is an organisation dedicated to the advancement of European research and development in information technology and applied mathematics. Its member institutions aim to foster collaborative work within the European research community and to increase co-operation with European industry.

ERCIM is the European Host of the World Wide Web Consortium.

Consiglio Nazionale delle Ricerche
Area della Ricerca CNR di Pisa
Via G. Moruzzi 1, 56124 Pisa, Italy
www.iit.cnr.it

Centrum Wiskunde & Informatica
Science Park 123,
NL-1098 XG Amsterdam, The Netherlands
www.cwi.nl

Fonds National de la Recherche
6, rue Antoine de Saint-Exupéry, B.P. 1777
L-1017 Luxembourg-Kirchberg
www.fnr.lu

Foundation for Research and Technology – Hellas
Institute of Computer Science
P.O. Box 1385, GR-71110 Heraklion, Crete, Greece
www.ics.forth.gr

Fraunhofer ICT Group
Anna-Louisa-Karsch-Str. 2
10178 Berlin, Germany
www.iuk.fraunhofer.de

INESC
c/o INESC Porto, Campus da FEUP,
Rua Dr. Roberto Frias, nº 378,
4200-465 Porto, Portugal
www.inesc.pt

Institut National de Recherche en Informatique
et en Automatique
B.P. 105, F-78153 Le Chesnay, France
www.inria.fr

I.S.I. – Industrial Systems Institute
Patras Science Park building
Platani, Patras, Greece, GR-26504
www.isi.gr

Norwegian University of Science and Technology
Faculty of Information Technology, Mathematics and Electrical Engineering, N 7491 Trondheim, Norway
http://www.ntnu.no/

RISE SICS
Box 1263,
SE-164 29 Kista, Sweden
http://www.sics.se/

SBA Research gGmbH
Floragasse 7, 1040 Wien, Austria
www.sba-research.org/

SIMULA
PO Box 134
1325 Lysaker, Norway
www.simula.no

Magyar Tudományos Akadémia
Számítástechnikai és Automatizálási Kutató Intézet
P.O. Box 63, H-1518 Budapest, Hungary
www.sztaki.hu/

TNO
PO Box 96829
2509 JE DEN HAAG
www.tno.nl

University of Cyprus
P.O. Box 20537
1678 Nicosia, Cyprus
www.cs.ucy.ac.cy/

Universty of Warsaw
Faculty of Mathematics, Informatics and Mechanics
Banacha 2, 02-097 Warsaw, Poland
www.mimuw.edu.pl/

VTT Technical Research Centre of Finland Ltd
PO Box 1000
FIN-02044 VTT, Finland
www.vttresearch.com